G-CSF ANALOG COMPOSITIONS AND METHODS

Cross Reference to Related Applications

The present application is a divisional of U.S. Patent Application No. 09/304,186 , filed May 3, 1999, which is a continuation of U.S. Patent Application No. 09/027,508, filed February 20, 1998, which is a continuation of U.S. Patent Application No. 08/956,812, filed October 23, 1997, which is a divisional of U.S. Patent Application No. 08/448,716, filed May 24, 1995, now Patent No. 5,790,421, which is a divisional of U.S. Patent Application No. 08/010,099, filed January 28, 1993, now Patent No. 5,581,476, which is hereby incorporated by reference.

Background of the Invention

Field of the Invention

This invention relates to granulocyte colony stimulating factor ("G-CSF") analogs, compositions containing such analogs, and related compositions. In another aspect, the present invention relates to nucleic acids encoding the present analogs or related nucleic acids, related host cells and vectors. In another aspect, the invention relates to computer programs and apparatuses for expressing the three dimensional structure of G-CSF and analogs thereof. In another aspect, the invention relates to methods for rationally designing G-CSF analogs and related compositions. In yet another aspect, the present invention relates to methods for treatment using the present G-CSF analogs.

Description of the Related Art

Hematopoiesis is controlled by two systems: the cells within the bone marrow microenvironment and growth factors. The growth factors, also called colony stimulating factors, stimulate committed progenitor cells to proliferate and to form colonies of differentiating blood cells. One of these factors is granulocyte colony stimulating factor, herein called G-CSF, which preferentially stimulates the growth and development of neutrophils, indicating a potential use in neutropenic states. Welte et al. *PNAS-USA* 82: 1526-1530 (1985); Souza et al. *Science* 232: 61-65 (1986) and Gabrilove, *J. Seminars in Hematology* 26:2 1-14 (1989).

In humans, endogenous G-CSF is detectable in blood plasma. Jones et al. *Bailliere's Clinical Hematology* 2:1 83-111 (1989). G-CSF is produced by fibroblasts, macrophages, T cells trophoblasts, endothelial cells and epithelial cells and is the expression product of a single copy gene comprised of four exons and five introns located on chromosome seventeen. Transcription of this locus produces a mRNA species which is differentially processed, resulting in two forms of G-CSF mRNA, one version coding for a protein of 177 amino acids, the other coding for a protein of 174 amino acids, Nagata et al. *EMBO J* 5: 575-581 (1986), and the form comprised of 174 amino acids has been found to have the greatest specific in vivo biological activity. G-CSF is species cross-reactive, such that when human G-CSF is administered to another mammal such as a mouse, canine or monkey, sustained neutrophil leukocytosis is elicited. Moore et al. *PNAS-USA* 84: 7134-7138 (1987).

Human G-CSF can be obtained and purified from a number of sources. Natural human G-CSF (nhG-CSF) can be isolated from the supernatants of cultured human tumor cell lines. The development of recombinant DNA technology, see, for instance, U.S. Patent 4,810,643 (Souza) incorporated herein by reference, has enabled the production of commercial scale quantities of G-CSF in glycosylated form as a product of eukaryotic host cell expression, and of G-CSF in non-glycosylated form as a product of prokaryotic host cell expression.

G-CSF has been found to be useful in the treatment of indications where an increase in neutrophils will provide benefits. For example, for cancer patients, G-CSF is beneficial as a means of selectively stimulating neutrophil production to compensate for hematopoietic deficits resulting from chemotherapy or radiation therapy. Other indications include treatment of various infectious diseases and related conditions, such as sepsis, which is typically caused by a metabolite of bacteria. G-CSF is also useful alone, or in combination with other compounds, such as other cytokines, for growth or expansion of cells in culture, for example, for bone marrow transplants.

Signal transduction, the way in which G-CSF effects cellular metabolism, is not currently thoroughly understood. G-CSF binds to a cell-surface receptor which apparently initiates the changes within particular progenitor cells, leading to cell differentiation.

Various altered G-CSF's have been reported. Generally, for design of drugs, certain changes are known to have certain structural effects. For example,

deleting one cysteine could result in the unfolding of a molecule which is, in its unaltered state, is normally folded via a disulfide bridge. There are other known methods for adding, deleting or substituting amino acids in order to change the function of a protein.

Recombinant human G-CSF mutants have been prepared, but the method of preparation does not include overall structure/function relationship information. For example, the mutation and biochemical modification of Cys^{18} has been reported. Kuga et al. *Biochem. Biophy. Res. Comm* 159: 103-111 (1989); Lu et al. *Arch. Biochem. Biophys.* 268: 81-92 (1989).

In U.S. Patent No. 4,810,643, entitled, "Production of Pluripotent Granulocyte Colony-Stimulating Factor" (as cited above), polypeptide analogs and peptide fragments of G-CSF are disclosed generally. Specific G-CSF analogs disclosed include those with the cysteins at positions 17, 36, 42, 64, and 74 (of the 174 amino acid species or of those having 175 amino acids, the additional amino acid being an N-terminal methionine) substituted with another amino acid, (such as serine), and G-CSF with an alanine in the first (N-terminal) position.

EP 0 335 423 entitled "Modified human G-CSF" reportedly discloses the modification of at least one amino group in a polypeptide having hG-CSF activity.

EP 0 272 703 entitled "Novel Polypeptide" reportedly discloses G-CSF derivatives having an amino acid substituted or deleted at or "in the neighborhood" of the N-terminus.

EP 0 459 630, entitled "Polypeptides" reportedly discloses derivatives of naturally occurring

G-CSF having at least one of the biological properties of naturally occurring G-CSF and a solution stability of at least 35% at 5mg/ml in which the derivative has at least Cys^{17} of the native sequence replaced by a Ser^{17} residue and Asp^{27} of the native sequence replaced by a Ser^{27} residue.

EP 0 256 843 entitled "Expression of G-CSF and Muteins Thereof and Their Uses" reportedly discloses a modified DNA sequence encoding G-CSF wherein the N-terminus is modified for enhanced expression of protein in recombinant host cells, without changing the amino acid sequence of the protein.

EP 0 243 153 entitled "Human G-CSF Protein Expression" reportedly discloses G-CSF to be modified by inactivating at least one yeast KEX2 protease processing site for increased yield in recombinant production using yeast.

Shaw, U.S. Patent No. 4,904,584, entitled "Site-Specific Homogeneous Modification of Polypeptides," reportedly discloses lysine altered proteins.

WO/9012874 reportedly discloses cysteine altered variants of proteins.

Australian patent application Document No. AU-A-10948/92, entitled, "Improved Activation of Recombinant Proteins" reportedly discloses the addition of amino acids to either terminus of a G-CSF molecule for the purpose of aiding in the folding of the molecule after prokaryotic expression.

Australian patent application Document No. AU-A-76380/91, entitled, "Muteins of the Granulocyte Colony Stimulating Factor (G-CSF)" reportedly discloses

muteins of the granulocyte stimulating factor G-CSF in the sequence Leu-Gly-His-Ser-Leu-Gly-Ile at position 50-56 of G-CSF with 174 amino acids, and position 53 to 59 of the G-CSF with 177 amino acids, or/and at least one of the four histadine residues at positions 43, 79, 156 and 170 of the mature G-CSF with 174 amino acids or at positions 46, 82, 159, or 173 of the mature G-CSF with 177 amino acids.

GB 2 213 821, entitled "Synthetic Human Granulocyte Colony Stimulating Factor Gene" reportedly discloses a synthetic G-CSF-encoding nucleic acid sequence incorporating restriction sites to facilitate the cassette mutagenesis of selected regions, and flanking restriction sites to facilitate the incorporation of the gene into a desired expression system.

G-CSF has reportedly been crystallized to some extent, i.e., EP 344 796, and the overall structure of G-CSF has been surmised, but only on a gross level. Bazan, *Immunology Today* 11: 350-354 (1990); Parry et al. *J.Molecular Recognition* 8: 107-110 (1988). To date, there have been no reports of the overall structure of G-CSF, and no systematic studies of the relationship of the overall structure and function of the molecule, studies which are essential to the systematic design of G-CSF analogs. Accordingly, there exists a need for a method of this systematic design of G-CSF analogs, and the resultant compositions.

Summary of the Invention

The three dimensional structure of G-CSF has now been determined to the atomic level. From this

three-dimensional structure, one can now forecast with substantial certainty how changes in the composition of a G-CSF molecule may result in structural changes. These structural characteristics may be correlated with biological activity to design and produce G-CSF analogs.

Although others had speculated regarding the three dimensional structure of G-CSF, Bazan, *Immunology Today* 11: 350-354 (1990); Parry et al. *J. Molecular Recognition* 8: 107-110 (1988), these speculations were of no help to those wishing to prepare G-CSF analogs either because the surmised structure was incorrect (Parry et al., supra) and/or because the surmised structure provided no detail correlating the constituent moieties with structure. The present determination of the three-dimensional structure to the atomic level is by far the most complete analysis to date, and provides important information to those wishing to design and prepare G-CSF analogs. For example, from the present three dimensional structural analysis, precise areas of hydrophobicity and hydrophilicity have been determined.

Relative hydrophobicity is important because it directly relates to the stability of the molecule. Generally, biological molecules, found in aqueous environments, are externally hydrophilic and internally hydrophobic; in accordance with the second law of thermodynamics provides, this is the lowest energy state and provides for stability. Although one could have speculated that G-CSF's internal core would be hydrophobic, and the outer areas would be hydrophilic, one would have had no way of knowing specific hydrophobic or hydrophilic areas. With the presently

provided knowledge of areas of hydrophobicity/-philicity, one may forecast with substantial certainty which changes to the G-CSF molecule will affect the overall structure of the molecule.

As a general rule, one may use knowledge of the geography of the hydrophobic and hydrophilic regions to design analogs in which the overall G-CSF structure is not changed, but change does affect biological activity ("biological activity" being used here in its broadest sense to denote function). One may correlate biological activity to structure. If the structure is not changed, and the mutation has no effect on biological activity, then the mutation has no biological function. If, however, the structure is not changed and the mutation does affect biological activity, then the residue (or atom) is essential to at least one biological function. Some of the present working examples were designed to provide no change in overall structure, yet have a change in biological function.

Based on the correlation of structure to biological activity, one aspect of the present invention relates to G-CSF analogs. These analogs are molecules which have more, fewer, different or modified amino acid residues from the G-CSF amino acid sequence. The modifications may be by addition, substitution, or deletion of one or more amino acid residues. The modification may include the addition or substitution of analogs of the amino acids themselves, such as peptidomimetics or amino acids with altered moieties such as altered side groups. The G-CSF used as a basis for comparison may be of human, animal or recombinant nucleic acid-technology origin (although the working

examples disclosed herein are based on the recombinant production of the 174 amino acid species of human G-CSF, having an extra N-terminus methionyl residue). The analogs may possess functions different from natural human G-CSF molecule, or may exhibit the same functions, or varying degrees of the same functions. For example, the analogs may be designed to have a higher or lower biological activity, have a longer shelf-life or a decrease in stability, be easier to formulate, or more difficult to combine with other ingredients. The analogs may have no hematopoietic activity, and may therefore be useful as an antagonist against G-CSF effect (as, for example, in the overproduction of G-CSF). From time to time herein the present analogs are referred to as proteins or peptides for convenience, but contemplated herein are other types of molecules, such as peptidomimetics or chemically modified peptides.

In another aspect, the present invention relates to related compositions containing a G-CSF analog as an active ingredient. The term, "related composition," as used herein, is meant to denote a composition which may be obtained once the identity of the G-CSF analog is ascertained (such as a G-CSF analog labeled with a detectable label, related receptor or pharmaceutical composition). Also considered a related composition are chemically modified versions of the G-CSF analog, such as those having attached at least one polyethylene glycol molecule.

For example, one may prepare a G-CSF analog to which a detectable label is attached, such as a fluorescent, chemiluminescent or radioactive molecule.

Another example is a pharmaceutical composition which may be formulated by known techniques using known materials, see, Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA 18042) pp. 1435-1712, which are herein incorporated by reference. Generally, the formulation will depend on a variety of factors such as administration, stability, production concerns and other factors. The G-CSF analog may be administered by injection or by pulmonary administration via inhalation. Enteric dosage forms may also be available for the present G-CSF analog compositions, and therefore oral administration may be effective. G-CSF analogs may be inserted into liposomes or other microcarriers for delivery, and may be formulated in gels or other compositions for sustained release. Although preferred compositions will vary depending on the use to which the composition will be put, generally, for G-CSF analogs having at least one of the biological activities of natural G-CSF, preferred pharmaceutical compositions are those prepared for subcutaneous injection or for pulmonary administration via inhalation, although the particular formulations for each type of administration will depend on the characteristics of the analog.

Another example of related composition is a receptor for the present analog. As used herein, the term "receptor" indicates a moiety which selectively binds to the present analog molecule. For example, antibodies, or fragments thereof, or "recombinant antibodies" (see Huse et al. *Science* 246: 1275 (1989)) may be used as receptors. Selective binding does not mean only specific binding (although binding-specific

receptors are encompassed herein), but rather that the binding is not a random event. Receptors may be on the cell surface or intra- or extra-cellular, and may act to effectuate, inhibit or localize the biological activity of the present analogs. Receptor binding may also be a triggering mechanism for a cascade of activity indirectly related to the analog itself. Also contemplated herein are nucleic acids, vectors containing such nucleic acids and host cells containing such nucleic acids which encode such receptors.

Another example of a related composition is a G-CSF analog with a chemical moiety attached. Generally, chemical modification may alter biological activity or antigenicity of a protein, or may alter other characteristics, and these factors will be taken into account by a skilled practitioner. As noted above, one example of such chemical moiety is polyethylene glycol. Modification may include the addition of one or more hydrophilic or hydrophobic polymer molecules, fatty acid molecules, or polysaccharide molecules. Examples of chemical modifiers include polyethylene glycol, alklpolyethylene glycols, DI-poly(amino acids), polyvinylpyrrolidone, polyvinyl alcohol, pyran copolymer, acetic acid/acylation, proprionic acid, palmitic acid, stearic acid, dextran, carboxymethyl cellulose, pullulan, or agarose. See, Francis, *Focus on Growth Factors* 3: 4-10 (May 1992)(published by Mediscript, Mountview Court, Friern Barnet Lane, London N20 OLD, UK). Also, chemical modification may include an additional protein or portion thereof, use of a cytotoxic agent, or an antibody. The chemical modification may also include lecithin.

In another aspect, the present invention relates to nucleic acids encoding such analogs. The nucleic acids may be DNAs or RNAs or derivatives thereof, and will typically be cloned and expressed on a vector, such as a phage or plasmid containing appropriate regulatory sequences. The nucleic acids may be labeled (such as using a radioactive, chemiluminescent, or fluorescent label) for diagnostic or prognostic purposes, for example. The nucleic acid sequence may be optimized for expression, such as including codons preferred for bacterial expression. The nucleic acid and its complementary strand, and modifications thereof which do not prevent encoding of the desired analog are here contemplated.

In another aspect, the present invention relates to host cells containing the above nucleic acids encoding the present analogs. Host cells may be eukaryotic or prokaryotic, and expression systems may include extra steps relating to the attachment (or prevention) of sugar groups (glycosylation), proper folding of the molecule, the addition or deletion of leader sequences or other factors incident to recombinant expression.

In another aspect the present invention relates to antisense nucleic acids which act to prevent or modify the type or amount of expression of such nucleic acid sequences. These may be prepared by known methods.

In another aspect of the present invention, the nucleic acids encoding a present analog may be used for gene therapy purposes, for example, by placing a vector containing the analog-encoding sequence into a recipient so the nucleic acid itself is expressed

inside the recipient who is in need of the analog composition. The vector may first be placed in a carrier, such as a cell, and then the carrier placed into the recipient. Such expression may be localized or systemic. Other carriers include non-naturally occurring carriers, such as liposomes or other microcarriers or particles, which may act to mediate gene transfer into a recipient.

The present invention also provides for computer programs for the expression (such as visual display) of the G-CSF or analog three dimensional structure, and further, a computer program which expresses the identity of each constituent of a G-CSF molecule and the precise location within the overall structure of that constituent, down to the atomic level. Set forth below is one example of such program. There are many currently available computer programs for the expression of the three dimensional structure of a molecule. Generally, these programs provide for inputting of the coordinates for the three dimensional structure of a molecule (i.e., for example, a numerical assignment for each atom of a G-CSF molecule along an x, y, and z axis), means to express (such as visually display) such coordinates, means to alter such coordinates and means to express an image of a molecule having such altered coordinates. One may program crystallographic information, i.e., the coordinates of the location of the atoms of a G-CSF molecule in three dimension space, wherein such coordinates have been obtained from crystallographic analysis of said G-CSF molecule, into such programs to generate a computer program for the expression (such as visual display) of the G-CSF three dimensional structure. Also provided,

therefore, is a computer program for the expression of G-CSF analog three dimensional structure. Preferred is the computer program Insight II, version 4, available from Biosym, San Diego, California, with the coordinates as set forth in FIGURE 5 input. Preferred expression means is on a Silicon Graphics 320 VGX computer, with Crystal Eyes glasses (also available from Silicon Graphics), which allows one to view the G-CSF molecule or its analog stereoscopically. Alternatively, the present G-CSF crystallographic coordinates and diffraction data are also deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 19723, USA. One may use these data in preparing a different computer program for expression of the three dimensional structure of a G-CSF molecule or analog thereof. Therefore, another aspect of the present invention is a computer program for the expression of the three dimensional structure of a G-CSF molecule. Also provided is said computer program for visual display of the three dimensional structure of a G-CSF molecule; and further, said program having means for altering such visual display. Apparatus useful for expression of such computer program, particularly for the visual display of the computer image of said three dimensional structure of a G-CSF molecule or analog thereof is also therefore here provided, as well as means for preparing said computer program and apparatus.

The computer program is useful for preparation of G-CSF analogs because one may select specific sites on the G-CSF molecule for alteration and readily ascertain the effect the alteration will have on the overall structure of the G-CSF molecule.

Selection of said site for alteration will depend on the desired biological characteristic of the G-CSF analog. If one were to randomly change said G-CSF molecule (r-met-hu-G-CSF) there would be 175^{20} possible substitutions, and even more analogs having multiple changes, additions or deletions. By viewing the three dimensional structure wherein said structure is correlated with the composition of the molecule, the selection for sites of alteration is no longer a random event, but sites for alteration may be determined rationally.

As set forth above, identity of the three dimensional structure of G-CSF, including the placement of each constituent down to the atomic level has now yielded information regarding which moieties are necessary to maintain the overall structure of the G-CSF molecule. One may therefore select whether to maintain the overall structure of the G-CSF molecule when preparing a G-CSF analog of the present invention, or whether (and how) to change the overall structure of the G-CSF molecule when preparing a G-CSF analog of the present invention. Optionally, once one has prepared such analog, one may test such analog for a desired characteristic.

One may, for example, seek to maintain the overall structure possessed by a non-altered natural or recombinant G-CSF molecule. The overall structure is presented in Figures 2, 3, and 4, and is described in more detail below. Maintenance of the overall structure may ensure receptor binding, a necessary characteristic for an analog possessing the hematopoietic capabilities of natural G-CSF (if no receptor binding, signal transduction does not result

from the presence of the analog). It is contemplated that one class of G-CSF analogs will possess the three dimensional core structure of a natural or recombinant (non-altered) G-CSF molecule, yet possess different characteristics, such as an increased ability to selectively stimulate neutrophils. Another class of G-CSF analogs are those with a different overall structure which diminishes the ability of a G-CSF analog molecule to bind to a G-CSF receptor, and possesses a diminished ability to selectively stimulate neutrophils as compared to non-altered natural or recombinant G-CSF.

For example, it is now known which moieties within the internal regions of the G-CSF molecule are hydrophobic, and, correspondingly, which moieties on the external portion of the G-CSF molecule are hydrophilic. Without knowledge of the overall three dimensional structure, preferably to the atomic level as provided herein, one could not forecast which alterations within this hydrophobic internal area would result in a change in the overall structural conformation of the molecule. An overall structural change could result in a functional change, such as lack of receptor binding, for example, and therefore, diminishment of biological activity as found in non-altered G-CSF. Another class of G-CSF analogs is therefore G-CSF analogs which possess the same hydrophobicity as (non-altered) natural or recombinant G-CSF. More particularly, another class of G-CSF analogs possesses the same hydrophobic moieties within the four helical bundle of its internal core as those hydrophobic moieties possessed by (non-altered) natural

or recombinant G-CSF yet have a composition different from said non-altered natural or recombinant G-CSF.

Another example relates to external loops which are structures which connect the internal core (helices) of the G-CSF molecule. From the three dimensional structure -- including information regarding the spatial location of the amino acid residues -- one may forecast that certain changes in certain loops will not result in overall conformational changes. Therefore, another class of G-CSF analogs provided herein is that having an altered external loop but possessing the same overall structure as (non-altered) natural or recombinant G-CSF. More particularly, another class of G-CSF analogs provided herein are those having an altered external loop, said loop being selected from the loop present between helices A and B; between helices B and C; between helices C and D; between helices D and A, as those loops and helices are identified herein. More particularly, said loops, preferably the AB loop and/or the CD loop are altered to increase the half life of the molecule by stabilizing said loops. Such stabilization may be by connecting all or a portion of said loop(s) to a portion of an alpha helical bundle found in the core of a G-CSF (or analog) molecule. Such connection may be via beta sheet, salt bridge, disulfide bonds, hydrophobic interaction or other connecting means available to those skilled in the art, wherein such connecting means serves to stabilize said external loop or loops. For example, one may stabilize the AB or CD loops by connecting the AB loop to one of the helices within the internal region of the molecule.

The N-terminus also may be altered without change in the overall structure of a G-CSF molecule, because the N-terminus does not effect structural stability of the internal helices, and, although the external loops are preferred for modification, the same general statements apply to the N-terminus.

Additionally, such external loops may be the site(s) for chemical modification because in (non-altered) natural or recombinant G-CSF such loops are relatively flexible and tend not to interfere with receptor binding. Thus, there would be additional room for a chemical moiety to be directly attached (or indirectly attached via another chemical moiety which serves as a chemical connecting means). The chemical moiety may be selected from a variety of moieties available for modification of one or more function of a G-CSF molecule. For example, an external loop may provide sites for the addition of one or more polymer which serves to increase serum half-life, such as a polyethylene glycol molecule. Such polyethylene glycol molecule(s) may be added wherein said loop is altered to include additional lysines which have reactive side groups to which polyethylene glycol moieties are capable of attaching. Other classes of chemical moieties may also be attached to one or more external loops, including but not limited to other biologically active molecules, such as receptors, other therapeutic proteins (such as other hematopoietic factors which would engender a hybrid molecule), or cytotoxic agents (such as diphtheria toxin). This list is of course not complete; one skilled in the art possessed of the desired chemical moiety will have the means to effect attachment of said desired moiety to the desired

external loop. Therefore, another class of the present G-CSF analogs includes those with at least one alteration in an external loop wherein said alteration provides for the addition of a chemical moiety such as at least one polyethylene glycol molecule.

Deletions, such as deletions of sites recognized by proteins for degradation of the molecule, may also be effectual in the external loops. This provides alternative means for increasing half-life of a molecule otherwise having the G-CSF receptor binding and signal transduction capabilities (i.e., the ability to selectively stimulate the maturation of neutrophils). Therefore, another class of the present G-CSF analogs includes those with at least one alteration in an external loop wherein said alteration decreases the turnover of said analog by proteases. Preferred loops for such alterations are the AB loop and the CD loop. One may prepare an abbreviated G-CSF molecule by deleting a portion of the amino acid residues found in the external loops (identified in more detail below), said abbreviated G-CSF molecule may have additional advantages in preparation or in biological function.

Another example relates to the relative charges between amino acid residues which are in proximity to each other. As noted above, the G-CSF molecule contains a relatively tightly packed four helical bundle. Some of the faces on the helices face other helices. At the point (such as a residue) where a helix faces another helix, the two amino acid moieties which face each other may have the same charge, and thus tend to repel each other, which lends instability to the overall molecule. This may be

eliminated by changing the charge (to an opposite charge or a neutral charge) of one or both of the amino acid moieties so that there is no repelling. Therefore, another class of G-CSF analogs includes those G-CSF analogs having been altered to modify instability due to surface interactions, such as electron charge location.

In another aspect, the present invention relates to methods for designing G-CSF analogs and related compositions and the products of those methods. The end products of the methods may be the G-CSF analogs as defined above or related compositions. For instance, the examples disclosed herein demonstrate (a) the effects of changes in the constituents (i.e., chemical moieties) of the G-CSF molecule on the G-CSF structure, and (b) the effects of changes in structure on biological function. Essentially, therefore, another aspect of the present invention is a method for preparing a G-CSF analog comprising the steps of:

(a) viewing information conveying the three dimensional structure of a G-CSF molecule wherein the chemical moieties, such as each amino acid residue or each atom of each amino acid residue, of the G-CSF molecule are correlated with said structure,

(b) selecting from said information a site on a G-CSF molecule for alteration;

(c) preparing a G-CSF analog molecule having such alteration; and

(d) optionally, testing such G-CSF analog molecule for a desired characteristic.

One may use the here provided computer programs for a computer-based method for preparing a G-CSF analog. Another aspect of the present invention

is therefore a computer based method for preparing a G-CSF analog comprising the steps of:

(a) providing computer expression of the three dimensional structure of a G-CSF molecule wherein the chemical moieties, such as each amino acid residue or each atom of each amino acid residue, of the G-CSF molecule are correlated with said structure;

(b) selecting from said computer expression a site on a G-CSF molecule for alteration;

(c) preparing a G-CSF molecule having such alteration; and,

(d) optionally, testing such G-CSF molecule for a desired characteristic.

More specifically, the present invention provides a method for preparing a G-CSF analog comprising the steps of:

(a) viewing the three dimensional structure of a G-CSF molecule via a computer, said computer programmed (i) to express the coordinates of a G-CSF molecule in three dimensional space, and (ii) to allow for entry of information for alteration of said G-CSF expression and viewing thereof;

(b) selecting a site on said visual image of said G-CSF molecule for alteration;

(c) entering information for said alteration on said computer;

(d) viewing a three dimensional structure of said altered G-CSF molecule via said computer;

(e) optionally repeating steps (a)-(e);

(f) preparing a G-CSF analog with said alteration; and

(g) optionally testing said G-CSF analog for a desired characteristic.

In another aspect, the present invention relates to methods of using the present G-CSF analogs and related compositions and methods for the treatment or protection of mammals, either alone or in combination with other hematopoietic factors or drugs in the treatment of hematopoietic disorders. It is contemplated that one aspect of designing G-CSF analogs will be the goal of enhancing or modifying the characteristics non-modified G-CSF is known to have. For example, the present analogs may possess enhanced or modified activities, so, where G-CSF is useful in the treatment of (for example) neutropenia, the present compositions and methods may also be of such use.

Another example is the modification of G-CSF for the purpose of interacting more effectively when used in combination with other factors particularly in the treatment of hematopoietic disorders. One example of such combination use is to use an early-acting hematopoietic factor (i.e., a factor which acts earlier in the hematopoiesis cascade on relatively undifferentiated cells) and either simultaneously or in seriatim use of a later-acting hematopoietic factor, such as G-CSF or analog thereof (as G-CSF acts on the CFU-GM lineage in the selective stimulation of neutrophils). The present methods and compositions may be useful in therapy involving such combinations or "cocktails" of hematopoietic factors.

The present compositions and methods may also be useful in the treatment of leukopenia, mylogenous leukemia, severe chronic neutropenia, aplastic anemia, glycogen storage disease, mucosistitis, and other bone marrow failure states. The present compositions and methods may also be useful in the treatment of

hematopoietic deficits arising from chemotherapy or from radiation therapy. The success of bone marrow transplantation, or the use of peripheral blood progenitor cells for transplantation, for example, may be enhanced by application of the present compositions (proteins or nucleic acids for gene therapy) and methods. The present compositions and methods may also be useful in the treatment of infectious diseases, such in the context of wound healing, burn treatment, bacteremia, septicemia, fungal infections, endocarditis, osteopyelitis, infection related to abdominal trauma, infections not responding to antibiotics, pneumonia and the treatment of bacterial inflammation may also benefit from the application of the present compositions and methods. In addition, the present compositions and methods may be useful in the treatment of leukemia based upon a reported ability to differentiate leukemic cells. Welte et al. *PNAS-USA* 82: 1526-1530 (1985). Other applications include the treatment of individuals with tumors, using the present compositions and methods, optionally in the presence of receptors (such as antibodies) which bind to the tumor cells. For review articles on therapeutic applications, see Lieshhke et al. *N.Engl.J.Med.* 327: 28-34, 99-106 (1992) both of which are herein incorporated by reference.

The present compositions and methods may also be useful to act as intermediaries in the production of other moieties; for example, G-CSF has been reported to influence the production of other hematopoietic factors and this function (if ascertained) may be enhanced or modified via the present compositions and/or methods.

The compositions related to the present G-CSF analogs, such as receptors, may be useful to act as an antagonist which prevents the activity of G-CSF or an analog. One may obtain a composition with some or all of the activity of non-altered G-CSF or a G-CSF analog, and add one or more chemical moieties to alter one or more properties of such G-CSF or analog. With knowledge of the three dimensional conformation, one may forecast the best geographic location for such chemical modification to achieve the desired effect.

General objectives in chemical modification may include improved half-life (such as reduced renal, immunological or cellular clearance), altered bioactivity (such as altered enzymatic properties, dissociated bioactivities or activity in organic solvents), reduced toxicity (such as concealing toxic epitopes, compartmentalization, and selective biodistribution), altered immunoreactivity (reduced immunogenicity, reduced antigenicity or adjuvant action), or altered physical properties (such as increased solubility, improved thermal stability, improved mechanical stability, or conformational stabilization). See Francis, *Focus on Growth Factors* 3: 4-10 (May 1992)(published by Mediscript, Mountview Court, Friern Barnet Lane, London N20 OLD, UK).

The examples below are illustrative of the present invention and are not intended as a limitation. It is understood that variations and modifications will occur to those skilled in the art, and it is intended that the appended claims cover all such equivalent variations which come within the scope of the invention as claimed.

Brief Description of the Drawings

FIGURE 1 is an illustration of the amino acid sequence of the 174 amino acid species of G-CSF with an additional N-terminal methionine (Seq. ID. No. 2).

FIGURES 2A-2G are a topology diagrams of the crystalline structure of G-CSF, as well as hGH, pGH, GM-CSF, INF-ß, IL-2, and IL-4. These illustrations are based on inspection of cited references. The length of secondary structural elements are drawn in proportion to the number of residues. A, B, C, and D helices are labeled according to the scheme used herein for G-CSF. For INF-ß, the original labeling of helices is indicated in parentheses.

FIGURE 3 is an "ribbon diagram" of the three dimensional structure of G-CSF. Helix A is amino acid residues 11-39 (numbered according to FIGURE 1, above Seq. ID. No. 2), helix B is amino acid residues 72-91, helix C is amino acid residues 100-123, and helix D is amino acid residues 143-173. The relatively short 3^{10} helix is at amino acid residues 45-48, and the alpha helix is at amino acid residues 48-53. Residues 93-95 form almost one turn of a left handed helix.

FIGURE 4 is a "barrel diagram" of the three dimensional structure of G-CSF. Shown in various shades of gray are the overall cylinders and their orientations for the three dimensional structure of G-CSF. The numbers indicate amino acid residue position according to FIGURE 1 (Seq. ID. No. 2) above.

FIGURES 5 (1-41) is a list of the coordinates used to generate a computer-aided visual image of the three dimensional structure of G-CSF. The coordinates are set forth below. The columns correspond to separate field:

(i) Field 1 (from the left hand side) is the atom,

(ii) Field 2 is the assigned atom number,

(iii) Field 3 is the atom name (according to the periodic table standard nomenclature, with CB being carbon atom Beta, CG is Carbon atom Gamma, etc.);

(iv) Field 4 is the residue type (according to three letter nomenclature for amino acids as found in , i.e., Stryer, Biochemistry, 3d Ed., W.H. Freeman & Co., New York 1988, inside back cover);

(v) Fields 5-7 are the x-axis, y-axis and z-axis positions of the atom;

(vi) Field 8 (often a "1.00") designates occupancy at that position;

(vii) Field 9 designates the B-factor;

(viii) Field 10 designates the molecule designation three molecules (designated a, b, and c) of G-CSF crystallized together as a unit. The designation a, b, or c indicates which coordinates are from which molecule. The number after the letter (1, 2, or 3) indicates the assigned amino acid residue position, with molecule A having assigned positions 10-175, molecule B having assigned positions 210-375, and molecule C having assigned positions 410-575. These positions were so designated so that there would be no overlap among the three molecules which crystallized together. (The "W" designation indicates water).

FIGURES 6A-6C are schematic representations of the strategy involved in refining the crystallization matrix for parameters involved in crystallization. The crystallization matrix corresponds to the final concentration of the components (salts, buffers and precipitants) of the

crystallization solutions in the wells of a 24 well tissue culture plate. These concentrations are produced by pipetting the appropriate volume of stock solutions into the wells of the microtiter plate. To design the matrix, the crystallographer decides on an upper and lower concentration of the component. These upper and lower concentrations can be pipetted along either the rows (i.e., A1-A6, B1-B6, C1-C6 or D1-D6) or along the entire tray (A1-D6). The former method is useful for checking reproducibility of crystal growth of a single component along a limited number of wells, whereas the later method is more useful in initial screening. The results of several stages of refinement of the crystallization matrix are illustrated by a representation of three plates. The increase in shading in the wells indicates a positive crystallization result which, in the final stages, would be X-ray quality crystals but in the initial stages could be oil droplets, granular precipitates or small crystals approximately less than 0.05mm in size. Part A represents an initial screen of one parameter in which the range of concentration between the first well (A1) and last well (D6) is large and the concentration increase between wells is calculated as ((concentration A1)-(concentration D6))/23). Part B represents that in later stages of the crystallization matrix refinement of the concentration spread between A1 and D6 would be reduced which would result in more crystals formed per plate. Part C indicates a final stage of matrix refinement in which quality crystals are found in most wells of the plate.

Detailed Description of the Invention

The present invention grows out of the discovery of the three dimensional structure of G-CSF. This three dimensional structure has been expressed via computer program for stereoscopic viewing. By viewing this stereoscopically, structure-function relationships identified and G-CSF analogs have been designed and made.

The Overall Three Dimensional Structure of G-CSF

The G-CSF used to ascertain the structure was a non-glycosylated 174 amino acid species having an extra N-terminal methionine residue incident to bacterial expression. The DNA (Seq. ID. No. 1) and amino acid sequence (Seq. ID. No. 2) of this G-CSF are illustrated in FIGURE 1.

Overall, the three dimensional structure of G-CSF is predominantly helical, with 103 of the 175 residues forming a 4-alpha-helical bundle. The only other secondary structure is found in the loop between the first two long helices where a 4 residue 3^{10} helix is immediately followed by a 6 residue alpha helix. As shown in FIGURE 2, the overall structure has been compared with the structure reported for other proteins: growth hormone (Abdel-Meguid et al. *PNAS-USA* 84: 6434 (1987) and Vos et al. *Science* 255: 305-312 (1992)), granulocyte macrophage colony stimulating factor (Diederichs et al. *Science* 254: 1779-1782 (1991)), interferon-ß (Senda et al. *EMBO J.* 11: 3193-3201 (1992)), interleukin-2 (McKay *Science* 257: 1673-1677 (1992)) and interleukin-4 (Powers et al. *Science* 256: 1673-1677 (1992), and Smith et al. *J. Mol. Biol.* 224: 899-904 (1992)). Structural similarity among

these growth factors occurs despite the absence of similarity in their amino acid sequences.

Presently, the structural information was correlation of G-CSF biochemistry, and this can be summarized as follows (with sequence position 1 being at the N-terminus):

TABLE 1

Sequence Position	Description of Structure	Analysis
1-10	Extended chain	Deletion causes no loss of biological activity
Cys^{18}	Partially buried	Reactive with DTNB and Thimersososl but not with iodo-acetate
34	Alternative splice site	Insertion reduces biological activity
20-47 (inclusive)	Helix A, first disulfide and portion of AB helix	Predicted receptor binding region based on neutralizing antibody data
20, 23, 24	Helix A	Single alanine mutation of residue(s) reduces biological activity. Predicted receptor binding (Site B).
165-175 (inclusive)	Carboxy terminus	Deletion reduces biological activity

This biochemical information, having been gleaned from antibody binding studies, see Layton et al. *Biochemistry* 266: 23815-23823 (1991), was superimposed on the three-dimensional structure in order to design G-CSF analogs . The design, preparation, and testing of these G-CSF analogs is described in Example 1 below.

Example 1:

This Example describes the preparation of crystalline G-CSF, the visualization of the three dimensional structure of recombinant human G-CSF via computer-generated image, the preparation of analogs, using site-directed mutagenesis or nucleic acid amplification methods, the biological assays and HPLC analysis used to analyze the G-CSF analogs, and the resulting determination of overall structure/function relationships. All cited publications are herein incorporated by reference.

A. Use of Automated Crystallization

The need for a three-dimensional structure of recombinant human granulocyte colony stimulating factor (r-hu-G-CSF), and the availability of large quantities of the purified protein, led to methods of crystal growth by incomplete factorial sampling and seeding. Starting with the implementation of incomplete factorial crystallization described by Jancarik et al. *J. Appl. Crystallogr.* 24: 409 (1991) solution conditions that yielded oil droplets and birefringence aggregates were ascertained. Also, software and hardware of an automated pipetting system were modified to produce some 400 different crystallization conditions per day. Weber *J. Appl. Crystallogr.* 20: 366-373 (1987). This procedure led to a crystallization solution which produced r-hu-G-CSF crystals.

The size, reproducibility and quality of the crystals was improved by a seeding method in which the number of "nucleation initiating units" was estimated by serial dilution of a seeding solution. These

methods yielded reproducible growth of 2.0mm r-hu-G-CSF crystals. The space group of these crystals is $P2_12_12_1$ with cell dimensions of a=90Å, b=110Å and c=49Å, and they diffract to a resolution of 2.0Å.

1. Overall Methodology

To search for the crystallizing conditions of a new protein, Carter et al. *J. Biol. Chem.* 254: 122219-12223 (1979) proposed the incomplete factorial method. They suggested that a sampling of a large number of randomly selected, but generally probable, crystallizing conditions may lead to a successful combination of reagents that produce protein crystallization. This idea was implemented by Jancarik et al. *J. Appl. Crystallogr.* 24: 409 (1991), who described 32 solutions for the initial crystallization trials which cover a range of pH, salts and precipitants. Here we describe an extension of their implementation to an expanded set of 70 solutions. To minimize the human effort and error of solution preparation, the method has been programmed for an automatic pipetting machine.

Following Weber's method of successive automated grid searching (SAGS), *J.Cryst. Growth* 90: 318-324 (1988), the robotic system was used to generate a series of solutions which continually refined the crystallization conditions of temperature, pH, salts and precipitant. Once a solution that could reproducibly grow crystals was determined, a seeding technique which greatly improved the quality of the crystals was developed. When these methods were combined, hundreds of diffraction quality crystals (crystals diffracting to at least about 2.5 Angstroms,

preferably having at least portions diffracting to below 2 Angstroms, and more preferably, approximately 1 Angstrom) were produced in a few days.

Generally, the method for crystallization, which may be used with any protein one desires to crystallize, comprises the steps of:

(a) combining aqueous aliquots of the desired protein with either (i) aliquots of a salt solution, each aliquot having a different concentration of salt; or (ii) aliquots of a precipitant solution, each aliquot having a different concentration of precipitant, optionally wherein each combined aliquot is combined in the presence of a range of pH;

(b) observing said combined aliquots for precrystalline formations, and selecting said salt or precipitant combination and said pH which is efficacious in producing precrystalline forms, or, if no precrystalline forms are so produced, increasing the protein starting concentration of said aqueous aliquots of protein;

(c) after said salt or said precipitant concentration is selected, repeating step (a) with said previously unselected solution in the presence of said selected concentration; and

(d) repeating step (b) and step (a) until a crystal of desired quality is obtained.

The above method may optionally be automated, which provides vast savings in time and labor. Preferred protein starting concentrations are between 10mg/ml and 20mg/ml, however this starting concentration will vary with the protein (the G-CSF below was analyzed using 33mg/ml). A preferred range of salt solution to begin analysis with is (NaCl) of

0-2.5M. A preferred precipitant is polyethylene glycol 8000, however, other precipitants include organic solvents (such as ethanol,), polyethylene glycol molecules having a molecular weight in the range of 500-20,000, and other precipitants known to those skilled in the art. The preferred pH range is pH 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. Precrystallization forms include oils, birefringement precipitants, small crystals (< approximately 0.05mm), medium crystals (approximately 0.5 to .5mm) and large crystals (> approximately 0.5mm). The preferred time for waiting to see a crystalline structure is 48 hours, although weekly observation is also preferred, and generally, after about one month, a different protein concentration is utilized (generally the protein concentration is increased). Automation is preferred, using the Accuflex system as modified. The preferred automation parameters are described below.

Generally, protein with a concentration between 10mg/ml and 20mg/ml was combined with a range of NaCl solutions from 0-2.5M, and each such combination was performed (separately) in the presence of the above range of concentrations. Once a precrystallization structure is observed, that salt concentration and pH range are optimized in a separate experiment, until the desired crystal quality is achieved. Next, the precipitant concentration, in the presence of varying levels of pH is also optimized. When both are optimized, the optimal conditions are performed at once to achieve the desired result (this is diagrammed in FIGURE 6).

a. Implementation of an Automated Pipetting System

Drops and reservoir solutions were prepared by an Accuflex pipetting system (ICN Pharmaceuticals, Costa Mesa, CA) which is controlled by a personal computer that sends ASCII codes through a standard serial interface. The pipetter samples six different solutions by means of a rotating valve and pipettes these solutions onto a plate whose translation in a x-y coordinate system can be controlled. The vertical component of the system manipulates a syringe that is capable both of dispensing and retrieving liquid.

The software provided with the Accuflex was based on the SAGS method as proposed by Cox et al. *J. Appl. Crystallogr.* 20: 366-373 (1987). This method involves the systematic variation of two major crystallization parameters, pH and precipitant concentration, with provision to vary two others. While building on these concepts, the software used here provided greater flexibility in the design and implementation of the crystallization solutions used in the automated grid searching strategy. As a result of this flexibility the present software also created a larger number of different solutions. This is essential for the implementation of the incomplete factorial method as described in that section below.

To improve the speed and design of the automated grid searching strategy, the Accuflex pipetting system required software and hardware modifications. The hardware changes allowed the use of two different micro-titer trays, one used for handing drop and one used for sitting drop experiments, and a Plexiglas tray which held 24 additional buffer, salt

and precipitant solutions. These additional solutions expanded the grid of crystallizing conditions that could be surveyed.

To utilize the hardware modifications, the pipetting software was written in two subroutines; one subroutine allows the crystallographer to design a matrix of crystallization solutions based on the concentrations of their components and the second subroutine to translate these concentrations into the computer code which pipettes the proper volumes of the solutions into the crystallization trays. The concentration matrices can be generated by either of two programs. The first program (MRF, available from Amgen Inc., Thousand Oaks, CA) refers to a list of stock solution concentrations supplied by the crystallographer and calculates the required volume to be pipette to achieve the designated concentration. The second method, which is preferred, incorporates a spread sheet program (Lotus™) which can be used to make more sophisticated gradients of precipitants or pH. The concentration matrix created by either program is interpreted by the control program (SUX, a modification of the program found in the Accuflex pipetter originally and available from Amgen Inc., Thousand Oaks, CA) and the wells are filled accordingly.

b. Implementation of the Incomplete Factorial Method

The convenience of the modified pipetting system for preparing diverse solutions improved the implementation of an expanded incomplete factorial method. The development of a new set of crystallization solutions having "random" components

was generated using the program INFAC, Carter et al. *J.Cryst. Growth* 90: 60-73(1988) which produced a list containing 96 random combinations of one factor from three variables. Combinations of calcium and phosphate which immediately precipitated were eliminated, leaving 70 distinct combinations of precipitants, salts and buffers. These combinations were prepared using the automated pipetter and incubated for one week. The mixtures were inspected and solutions which formed precipitants were prepared again with lower concentrations of their components. This was repeated until all wells were clear of precipitant.

c. Crystallization of r-hu-G-CSF

Several different crystallization strategies were used to find a solution which produced x-ray quality crystals. These strategies included the use of the incomplete factorial method, refinement of the crystallization conditions using successive automated grid searches (SAGS), implementation of a seeding technique and development of a crystal production procedure which yielded hundreds of quality crystals overnight. Unless otherwise noted the screening and production of r-hu-G-CSF crystals utilized the hanging drop vapor diffusion method. Afinsen et al. "Physical Principles Of Protein Crystallization." In: Eisenberg (ed.), *Advances in Protein Chemistry* 41: 1-33 (1991).

The initial screening for crystallization conditions of r-hu-G-CSF used the Jancarik et al. *J.Appl.Crystallogr.* 24: 409 (1991) incomplete factorial method which resulted in several solutions that produced "precrystallization" results. These results included birefringent precipitants, oils and very small

crystals (< .05mm). These precrystallizations solutions then served as the starting points for systematic screening.

The screening process required the development of crystallization matrices. These matrices corresponded to the concentration of the components in the crystallization solutions and were created using the IBM-PC based spread sheet Lotus™ and implemented with the modified Accuflex pipetting system. The strategy in designing the matrices was to vary one crystallization condition (such as salt concentration) while holding the other conditions such as pH, and precipitant concentration constant. At the start of screening, the concentration range of the varied condition was large but the concentration was successively refined until all wells in the micro-titer tray produced the same crystallization result. These results were scored as follows: crystals, birefringement precipitate, granular precipitate, oil droplets and amorphous mass. If the concentration of a crystallization parameter did not produce at least a precipitant, the concentration of that parameter was increased until a precipitant formed. After each tray was produced, it was left undisturbed for at least two days and then inspected for crystal growth. After this initial screening, the trays were then inspected on a weekly basis.

From this screening process, two independent solutions with the same pH and precipitant but differing in salts (MgCl, $LiSO_4$) were identified which produced small (0.1 x 0.05 x 0.05mm) crystals. Based on these results, a new series of concentration matrices were produced which varied MgCl with respect

to $LiSO_4$ while keeping the other crystallization parameters constant. This series of experiments resulted in identification of a solution which produced diffraction quality crystals (> approximately 0.5mm) in about three weeks. To find this crystallization growth solution (100mM Mes pH 5.8, 380mM $MgCl_2$, 220mM $LiSO_4$ and 8% PEG 8k) approximately 8,000 conditions had been screened which consumed about 300mg of protein.

The size of the crystals depended on the number of crystals forming per drop. Typically 3 to 5 crystals would be formed with average size of (1.0 x 0.7 x 0.7mm). Two morphologies which had an identical space group ($P2_12_12_1$) and unit cell dimensions a=90.2, b=110.2, c=49.5 were obtained depending on whether or not seeding (see below) was implemented. Without seeding, the r-hu-G-CSF crystals had one long flat surface and rounded edges.

When seeding was employed, crystals with sharp faces were observed in the drop within 4 to 6 hours (0.05 by 0.05 by 0.05mm). Within 24 hours, crystals had grown to (0.7 by 0.7 by 0.7mm) and continued to grow beyond 2mm depending on the number of crystals forming in the drop.

d. Seeding and Determination of Nucleation Initiation Sites.

The presently provided method for seeding crystals establishes the number of nucleation initiation units in each individual well used (here, after the optimum conditions for growing crystals had been determined). The method here is advantageous in that the number of "seeds" affects the quality of the crystals, and this in turn affects the degree of

resolution. The present seeding here also provides advantages in that with seeding, G-CSF crystal grows in a period of about three days, whereas without seeding, the growth takes approximately three weeks.

In one series of production growth (see methods), showers of small but well defined crystals were produced overnight (<0.01 x 0.01 x 0.01mm). Crystallization conditions were followed as described above except that a pipette tip employed in previously had been reused. Presumably, the crystal showering effect was caused by small nucleation units which had formed in the used tip and which provided sites of nucleation for the crystals. Addition of a small amount (0.5µl) of the drops containing the crystal showers to a new drop under standard production growth conditions resulted in a shower of crystals overnight. This method was used to produce several trays of drops containing crystal showers which we termed "seed stock".

The number of nucleation initiation units (NIU) contained within the "seed stock" drops was estimated to attempt to improve the reproducibility and quality of the r-hu-GCSF crystals. To determine the number of NIU in the "seed stock", an aliquot of the drop was serially diluted along a 96 well microtiter plate. The microtiter plate was prepared by adding 50µl of a solution containing equal volumes of r-hu-G-CSF (33mg/ml) and the crystal growth solution (described above) in each well. An aliquot (3µl) of one of the "seed stock" drops was transferred to the first well of the microtiter plate. The solution in the well was mixed and 3µl was then transferred to the next well along the row of the microtiter plate. Each

row of the microtiter plate was similarly prepared and the tray was sealed with plastic tape. Overnight, small crystals formed in the bottom of the wells of the microtiter plate and the number of crystals in the wells were correlated to the dilution of the original "seed stock". To produce large single crystals, the "seed stock" drop was appropriately diluted into fresh CGS and then an aliquot of this solution containing the NIU was transferred to a drop

Once crystallization conditions had been optimized, crystals were grown in a production method in which 3ml each of CGS and r-hu-G-CSF (33mg/ml) were mixed to create five trays (each having 24 wells). This method included the production of the refined crystallization solution in liter quantities, mixing this solution with protein and placing the protein/crystallization solution in either hanging drop or sitting drop trays. This process typically yielded 100 to 300 quality crystals (>0.5mm) in about five days.

e. Experimental Methods

Materials

Crystallographic information was obtained starting with r-hu-met-G-CSF with the amino acid sequence as provided in FIGURE 1 (Seq. ID. No. 2) with a specific activity of 1.0 +/- 0.6 x 10^8U/mg (as measured by cell mitogenesis assay in a 10mM acetate buffer at pH 4.0 (in Water for Injection) at a concentration of approximately 3mg/ml solution was concentrated with an Amicon concentrator at 75 psi using a YM10 filter. The solution was typically

concentrated 10 fold at 4°C and stored for several months.

Initial Screening

Crystals suitable for X-ray analysis were obtained by vapor-diffusion equilibrium using hanging drops. For preliminary screening, 7µl of the protein solution at 33mg/ml (as prepared above) was mixed with an equal volume of the well solution, placed on siliconized glass plates and suspended over the well solution utilizing Linbro tissue culture plates (Flow Laboratories, McLean, Va). All of the pipetting was performed with the Accuflex pipetter, however, trays were removed from the automated pipetter after the well solutions had been created and thoroughly mixed for at least ten minutes with a table top shaker. The Linbro trays were then returned to the pipetter which added the well and protein solutions to the siliconized cover slips. The cover slips were then inverted and sealed over 1ml of the well solutions with silicon grease.

The components of the automated crystallization system are as follows. A PC-DOS computer system was used to design a matrix of crystallization solutions based on the concentration of their components. These matrices were produced with either MRF of the Lotus™ spread sheet (described above). The final product of these programs is a data file. This file contains the information required by the SUX program to pipette the appropriate volume of the stock solutions to obtain the concentrations described in the matrices. The SUX program information was passed through a serial I/O port and used to dictate to the Accuflex pipetting system the position

of the valve relative to the stock solutions, the amount of solution to be retrieved, and then pipetted into the wells of the microtiter plates and the X-Y position of each well (the column/row of each well). Additional information was transmitted to the pipetter which included the Z position (height) of the syringe during filling as well as the position of a drain where the system pauses to purge the syringe between fillings of different solutions. The 24 well microtiter plate (either Linbro or Cryschem) and cover slip holder was placed on a plate which was moved in the X-Y plane. Movement of the plate allowed the pipetter to position the syringe to pipette into the wells. It also positioned the coverslips and vials and extract solutions from these sources. Prior the pipetting, the Linbro microtiter plates had a thin film of grease applied around the edges of the wells. After the crystallization solutions were prepared in the wells and before they were transferred to the cover slips, the microtiter plate was removed from the pipetting system, and solutions were allowed to mix on a table top shaker for ten minutes. After mixing, the well solution was either transferred to the cover slips (in the case of the hanging drop protocol) or transferred to the middle post in the well (in the case of the sitting drop protocol). Protein was extracted from a vial and added to the coverslip drop containing the well solution (or to the post). Plastic tape was applied to the top of the Cryschem plate to seal the wells.

Production Growth

Once conditions for crystallization had been optimized, crystal growth was performed utilizing a "production" method. The crystallization solution which contained 100mM Mes pH 5.8, 380mM MgCl2, 220mM LiSO4, and 8% PEG 8K was made in one liter quantities. Utilizing an Eppindorf syringe pipetter, 1ml aliquots of this solution were pipetted into each of the wells of the Linbro plate. A solution containing 50% of this solution and 50% G-CSF (33mg/ml) was mixed and pipetted onto the siliconized cover slips. Typical volumes of these drops were between 50 and 100µl and because of the large size of these drops, great care was taken in flipping the coverslips and suspending the drops over the wells.

Data Collection

The structure has been refined with X-PLOR (Bruniger, X-PLOR version 3.0, A system for crystallography and NMR, Yale University, New Haven, CT) against 2.2Å data collected on an R-AXIS (Molecular Structure, Corp. Houston, TX) imaging plate detector.

f. Observations

As an effective recombinant human therapeutic, r-hu-G-CSF has been produced in large quantities and gram levels have been made available for structural analysis. The crystallization methods provided herein are likely to find other applications as other proteins of interest become available. This method can be applied to any crystallographic project which has large quantities of protein (approximately >200mg). As one skilled in the art will recognize, the

present materials and methods may be modified and equivalent materials and methods may be available for crystallization of other proteins.

B. Computer Program for Visualizing the Three Dimensional Structure of G-CSF

Although diagrams, such as those in the Figures herein, are useful for visualizing the three dimensional structure of G-CSF, a computer program which allows for stereoscopic viewing of the molecule is contemplated as preferred. This stereoscopic viewing, or "virtual reality" as those in the art sometimes refer to it, allows one to visualize the structure in its three dimensional form from every angle in a wide range of resolution, from macromolecular structure down to the atomic level. The computer programs contemplated herein also allow one to change perspective of the viewing angle of the molecule, for example by rotating the molecule. The contemplated programs also respond to changes so that one may, for example, delete, add, or substitute one or more images of atoms, including entire amino acid residues, or add chemical moieties to existing or substituted groups, and visualize the change in structure.

Other computer based systems may be used; the elements being: (a) a means for entering information , such as orthogonal coordinates or other numerically assigned coordinates of the three dimensional structure of G-CSF; (b) a means for expressing such coordinates, such as visual means so that one may view the three dimensional structure and correlate such three dimensional structure with the composition of the G-CSF

molecule, such as the amino acid composition; and (c) optionally, means for entering information which alters the composition of the G-CSF molecule expressed, so that the image of such three dimensional structure displays the altered composition.

The coordinates for the preferred computer program used are presented in FIGURE 5. The preferred computer program is Insight II, version 4, available from Biosym in San Diego, CA. For the raw crystallographic structure, the observed intensities of the diffraction data ("F-obs") and the orthogonal coordinates are also deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 19723, USA and these are herein incorporated by reference.

Once the coordinates are entered into the Insight II program, one can easily display the three dimensional G-CSF molecule representation on a computer screen. The preferred computer system for display is Silicon Graphics 320 VGX (San Diego, CA). For stereoscopic viewing, one may wear eyewear (Crystal Eyes, Silicon Graphics) which allows one to visualize the G-CSF molecule in three dimensions stereoscopically, so one may turn the molecule and envision molecular design.

Thus, the present invention provides a method of designing or preparing a G-CSF analog with the aid of a computer comprising:

(a) providing said computer with the means for displaying the three dimensional structure of a G-CSF molecule including displaying the composition of moieties of said G-CSF molecule, preferably displaying the three dimensional location of each amino acid, and

more preferably displaying the three dimensional location of each atom of a G-CSF molecule;

(b) viewing said display;

(c) selecting a site on said display for alteration in the composition of said molecule or the location of a moiety; and

(d) preparing a G-CSF analog with such alteration.

The alteration may be selected based on the desired structural characteristics of the end-product G-CSF analog, and considerations for such design are described in more detail below. Such considerations include the location and compositions of hydrophobic amino acid residues, particularly residues internal to the helical structures of a G-CSF molecule which residues, when altered, alter the overall structure of the internal core of the molecule and may prevent receptor binding; the location and compositions of external loop structures, alteration of which may not affect the overall structure of the G-CSF molecule.

FIGURES 2-4 illustrate the overall three dimensional conformation in different ways. The topological diagram, the ribbon diagram, and the barrel diagram all illustrate aspects of the conformation of G-CSF.

FIGURE 2 illustrates a comparison between G-CSF and other molecules. There is a similarity of architecture, although these growth factors differ in the local conformations of their loops and bundle geometrics. The up-up-down-down topology with two long crossover connections is conserved, however, among all six of these molecules, despite the dissimilarity in amino acid sequence.

FIGURE 3 illustrates in more detail the secondary structure of recombinant human G-CSF. This ribbon diagram illustrates the handedness of the helices and their positions relative to each other.

FIGURE 4 illustrates in a different way the conformation of recombinant human G-CSF. This "barrel" diagram illustrates the overall architecture of recombinant human G-CSF.

C. Preparation of Analogs Using M13 Mutagenesis

This example relates to the preparation of G-CSF analogs using site directed mutagenesis techniques involving the single stranded bacteriophage M13, according to methods published in PCT Application No. WO 85/00817 (Souza et al., published February 28, 1985, herein incorporated by reference). This method essentially involves using a single-stranded nucleic acid template of the non-mutagenized sequence, and binding to it a smaller oligonucleotide containing the desired change in the sequence. Hybridization conditions allow for non-identical sequences to hybridize and the remaining sequence is filled in to be identical to the original template. What results is a double stranded molecule, with one of the two strands containing the desired change. This mutagenized single strand is separated, and used itself as a template for its complementary strand. This creates a double stranded molecule with the desired change.

The original G-CSF nucleic acid sequence used is presented in FIGURE 1 (Seq. ID. No. 1), and the oligonucleotides containing the mutagenized nucleic acid(s) are presented in Table 2. Abbreviations used

herein for amino acid residues and nucleotides are conventional, see Stryer, Biochemistry, 3d Ed., W.H. Freeman and Company, New York, N.Y. 1988, inside back cover.

The original G-CSF nucleic acid sequence was first placed into vector M13mp21. The DNA from single stranded phage M13mp21 containing the original G-CSF sequence was then isolated, and resuspended in water. For each reaction, 200ng of this DNA was mixed with a 1.5 pmole of phosphorylated oligonucleotide (Table 2) and suspended in 0.1M Tris, 0.01M $MgCl_2$, 0.005M DTT, 0.1mM ATP, pH 8.0. The DNAs were annealed by heating to 65°C and slowly cooling to room temperature.

Once cooled, 0.5mM of each ATP, dATP, dCTP, dGTP, TTP, one unit of T4 DNA ligase and one unit of Klenow fragment of E. coli polymerase 1 were added to the one unit of annealed DNA in 0.1M Tris, 0.025M NaCl, 0.01M $MgCl_2$, 0.01M DTT, pH 7.5.

The now double stranded, closed circular DNA was used to transfect E. coli without further purification. Plaques were screened by lifting the plaques with nitrocellulose filters, and then hybridizing the filters with single stranded DNA end-labeled with P^{32} for one hour at 55-60°. After hybridization, the filters were washed at 0-3°C below the melt temperature of the oligo (2°C for A-T, 4°C for G-C) which selectively left autoradiography signals corresponding to plaques with phage containing the mutated sequence. Positive clones were confirmed by sequencing.

Set forth below are the oligonucleotides used for each G-CSF analog prepared via the M13 mutagenesis

method. The nomenclature indicates the residue and the position of the original amino acid (i.e., lysine at position 17), and the residue and position of the substituted amino acid (i.e., arginine 17). A substitution involving more than one residue is indicated via superscript notation, with commas between the noted positions or a semicolon indicating different residues. Deletions with no substitutions are so noted. The oligonucleotide sequences used for M13-based mutagenesis are next indicated; these oligonucleotides were manufactured synthetically, although the method of preparation is not critical, any nucleic acid synthesis method and/or equipment may be used. The length of the oligo is also indicated. As indicated above, these oligos were allowed to contact the single stranded phage vector, and then single nucleotides were added to complete the G-CSF analog nucleic acid sequence.

Table 2

G-CSF ANALOGS	SEQUENCES(5'-> 3')	Length (nucleotide)	Seq.ID. Nos.
Lys^{17}->Arg^{17}	CTT TCT GCT GCG TTG TCT GGA ACA	24	Seq.ID. No.3
Lys^{24}->Arg^{24}	ACA GGT TCG TCG TAT CCA GGG TG	23	Seq.ID. No.4
Lys^{35}->Arg^{35}	CAC TGC AAG AAC GTC TGT GCG CT	23	Seq.ID. No.5
Lys^{41}->Arg^{41}	CGC TAC TTA CCG TCT GTG CCA TC	23	Seq.ID. No.6
$Lys^{17,24,35}$-> $Arg^{17,24,35}$	CTT TCT GCT GCG TTG TCT GGA ACA	24	Seq.ID. No.7
	ACA GGT TCG TCG TAT CCA GGG TG	23	Seq.ID. No.8
	CAC TGC AAG AAC GTC TGT GCG CT	23	Seq.ID. No.9
$Lys^{17,24,41}$-> $Arg^{17,24,41}$	CTT TCT GCT GCG TTG TCT GGA ACA	24	Seq.ID. No.10
	ACA GGT TCG TCG TAT CCA GGG TG	23	Seq.ID. No.11
	CGC TAC TTA CCG TCT GTC CCA TC	23	Seq.ID. No.12
$Lys^{17,35,41}$-> $Arg^{17,35,41}$	CTT TCT GCT GCG TTG TCT GGA ACA	24	Seq.ID. No.13
	CAC TGC AAG AAC GTC TGT GCG CT	23	Seq.ID. No.14
	CGC TAC TTA CCG TCT GTG CCA TC	23	Seq.ID. No.15

$Lys^{24,35,41}$->	ACA GGT TCG TCG TAT CCA GGG TG	23	Seq.ID. No.16
$Arg^{24,35,41}$	CAC TGC AAG AAC GTC TGT GCG CT	23	Seq.ID. No.17
	CGC TAC TTA CCG TCT GTG CCA TC	23	Seq.ID. No.18
$Lys^{17,24,35,41}$->	CTT TCT GCT GCG TTG TCT GGA ACA	24	Seq.ID. No.19
$Arg^{17,24,35,41}$	ACA GGT TCG TCG TAT CCA GGG TG	23	Seq.ID. No.20
	CAC TGC AAG AAC GTC TGT GCG CT	23	Seq.ID. No.21
	CGC TAC TTA CCG TCT GTG CCA TC	23	Seq.ID. No.22
Cys^{18}->Ala^{18}	TCT GCT GAA AGC TCT GGA ACA GG	23	Seq.ID. No.23
Gln^{68}->Glu^{68}	CTT GTC CAT CTG AAG CTC TTC AG	23	Seq.ID. No.24
$Cys^{37,43}$->	GAA AAA CTG TCC GCT ACT TAC AAA	37	Seq.ID. No.25
$Ser^{37,43}$	CTG TCC CAT CCG G		
Gln^{26}->Ala^{26}	TTC GTA AAA TCG CGG GTG ACG G	22	Seq.ID. No.26
Gln^{174}->Ala^{174}	TCA TCT GGC TGC GCC GTA ATA G	22	Seq.ID. No.27
Arg^{170}->Ala^{170}	CCG TGT TCT GGC TCA TCT GGC T	22	Seq.ID. No.28
Arg^{167}->Ala^{167}	GAA GTA TCT TAC GCT GTT CTG CGT	24	Seq.ID. No.29
Deletion 167	GAA GTA TCT TAC TAA GTT CTG CGT C	25	Seq.ID. No.30
Lys^{41}->Ala^{41}	CGC TAC TTA CGC ACT GTG CCA T	22	Seq.ID. No.31
His^{44}->Lys^{44}	CAA ACT GTG CAA GCC GGA AGA G	22	Seq.ID. No.32
Glu^{47}->Ala^{47}	CAT CCG GAA GCA CTG GTA CTG C	22	Seq.ID. No.33
Arg^{23}->Ala^{23}	GGA ACA GGT TGC TAA AAT CCA GG	23	Seq.ID. No.34
Lys^{24}->Ala^{24}	GAA CAG GTT CGT GCG ATC CAG GGT G	25	Seq.ID. No.35
Glu^{20}->Ala^{20}	GAA ATG TCT GGC ACA GGT TCG T	22	Seq.ID. No.36
Asp^{28}->Ala^{28}	TCC AGG GTG CCG GTG CTG C	19	Seq.ID. No.37
Met^{127}->Glu^{127}	AAG AGC TCG GTG AGG CAC CAG CT	23	Seq.ID. No.38
Met^{138}->Glu^{138}	CTC AAG GTG CTG AGC CGG CAT TC	23	Seq.ID. No.39
Met^{127}->Leu^{127}	GAG CTC GGT CTG GCA CCA GC	20	Seq.ID. No.40
Met^{138}->Leu^{138}	TCA AGG TGC TCT GCC GGC ATT	21	Seq.ID. No.41
Ser^{13}->Ala^{13}	TCT GCC GCA AGC CTT TCT GCT GA	23	Seq.ID. No.42
Lys^{17}->Ala^{17}	CTT TCT GCT GGC ATG TCT GGA ACA	24	Seq.ID. No.43
Gln^{121}->Ala^{121}	CTA TTT GGC AAG CGA TGG AAG AGC	24	Seq.ID. No.44
Glu^{124}->Ala^{124}	CAG ATG GAA GCG CTC GGT ATG	21	Seq.ID. No.45

$Met^{127,138}$-> $Leu^{127,138}$	GAG CTC GGT CTG GCA CCA GC TCA AGG TGC TCT GCC GGC ATT	20 21	Seq.ID. No.46 Seq.ID. No.47
**Glu^{20}->Ala^{20}; Ser^{13}->Gly^{13}	GAA ATG TCT GGC ACA GGT TCG T	22	Seq.ID. No.48

** This analog came about during the preparation of G-CSF analog Glu^{20}->Ala^{20}. As several clones were being sequenced to identify the Glu^{20}->Ala^{20} analog, the Glu^{20}->Ala^{20}; Ser^{13}->Gly^{13} analog was identified. This double mutant was the result of an in vitro Klenow DNA polymerase reaction mistake.

D. Preparation of G-CSF Analogs Using DNA Amplification

This example relates to methods for producing G-CSF analogs using a DNA amplification technique. Essentially, DNA encoding each analog was amplified in two separate pieces, combined, and then the total sequence itself amplified. Depending upon where the desired change in the original G-CSF DNA was to be made, internal primers were used to incorporate the change, and generate the two separate amplified pieces. For example, for amplification of the 5' end of the desired analog DNA, a 5' flanking primer (complementary to a sequence of the plasmid upstream from the G-CSF original DNA) was used at one end of the region to be amplified, and an internal primer, capable of hybridizing to the original DNA but incorporating the desired change, was used for priming the other end. The resulting amplified region stretched from the 5' flanking primer through the internal primer. The same was done for the 3' terminus, using a 3' flanking primer (complementary to a sequence of the plasmid downstream from the G-CSF original DNA) and an internal primer complementary to the region of the intended mutation. Once the two "halves" (which may or may not be equal in size, depending on the location of the internal primer) were amplified, the two "halves" were

allowed to connect. Once connected, the 5' flanking primer and the 3' flanking primer were used to amplify the entire sequence containing the desired change.

If more than one change is desired, the above process may be modified to incorporate the change into the internal primer, or the process may be repeated using a different internal primer. Alternatively, the gene amplification process may be used with other methods for creating changes in nucleic acid sequence, such as the phage based mutagenesis technique as described above. Examples of process for preparing analogs with more than one change are described below.

To create the G-CSF analogs described below, the template DNA used was the sequence as in FIGURE 1 plus certain flanking regions (from a plasmid containing the G-CSF coding region). These flanking regions were used as the 5' and 3' flanking primers and are set forth below. The amplification reactions were performed in 40µl volumes containing 10mM Tris-HCl, 1.5mM $MgCl_2$, 50mM KCl, 0.1mg/ml gelatin, pH 8.3 at 20°C. The 40µl reactions also contained 0.1mM of each dNTP, 10 pmoles of each primer, and 1ng of template DNA. Each amplification was repeated for 15 cycles. Each cycle consisted of 0.5 minutes at 94°C, 0.5 minutes at 50°C, and 0.75 minutes at 72°C. Flanking primers were 20 nucleotides in length and internal primers were 20 to 25 nucleotides in length. This resulted in multiple copies of double stranded DNA encoding either the front portion or the back portion of the desired G-CSF analog.

For combining the two "halves", $^1/_{40}$ of each of the two reactions was combined in a third DNA amplification reaction. The two portions were allowed

to anneal at the internal primer location, as their ends bearing the mutation were complementary, and following a cycle of polymerization, give rise to a full length DNA sequence. Once so annealed, the whole analog was amplified using the 5' and 3' flanking primers. This amplification process was repeated for 15 cycles as described above.

The completed, amplified analog DNA sequence was cleaved with XbaI and XhoI restriction endonuclease to produce cohesive ends for insertion into a vector. The cleaved DNA was placed into a plasmid vector, and that vector was used to transform E. coli. Transformants were challenged with kanamycin at 50ug/ml and incubated at 30°C. Production of G-CSF analog protein was confirmed by polyacrylamide gel electrophoresis of a whole cell lysate. The presence of the desired mutation was confirmed by DNA sequence analysis of plasmid purified from the production isolate. Cultures were then grown, and cells were harvested, and the G-CSF analogs were purified as set forth below.

Set forth below in Table 3 are the specific primers used for each analog made using gene amplification.

TABLE 3

Analog	Internal Primer(5'->3')	SEQ.ID.NO.
His^{44}->Ala^{44}	5'primer-TTCCGGAGCGCACAGTTTG	Seq.ID.No.49
	3'primer-CAAACTGTGGGCTCCGGAAGAGC	Seq.ID.No.50
Thr^{117}->Ala^{117}	5'primer-ATGCCAAATTGCAGTAGCAAAG	Seq.ID.No.51
	3'primer-CTTTGCTACTGCAATTTGGCAACA	Seq.ID.No.52
Asp^{110}->Ala^{110}	5'primer-ATCAGCTACTGCTAGCTGCAGA	Seq.ID.No.53
	3'primer-TCTGCAGCTAGCAGTAGCTGACT	Seq.ID.No.54
Gln^{21}->Ala^{21}	5'primer-TTACGAACCGCTTCCAGACATT	Seq.ID.No.55

	3'primer-AATGTCTGGAAGCGGTTCGTAAAAT	Seq.ID.No.56
Asp^{113}->Ala^{113}	5'primer-GTAGCAAATGCAGCTACATCTA	Seq.ID.No.57
	3'primer-TAGATGTAGCTGCATTTGCTACTAC	Seq.ID.No.58
His^{53}->Ala^{53}	5'primer-CCAAGAGAAGCACCCAGCAG	Seq.ID.No.59
	3'primer-CTGCTGGGTGCTTCTCTTGGGA	Seq.ID.No.60

For each analog, the following 5' flanking primer was used:

5'-CACTGGCGGTGATAATGAGC (Seq.ID. No.61)

For each analog, the following 3' flanking primer was used:

3'-GGTCATTACGGACCGGATC (Seq.ID. No.62)

1. Construction of Double Mutation

To make G-CSF analog $Gln^{12,21}$->$Glu^{12,21}$, two separate DNA amplifications were conducted to create the two DNA mutations. The template DNA used was the sequence as in FIGURE 1 (Seq.ID. No.1) plus certain flanking regions (from a plasmid containing the G-CSF coding region). The precise sequences are listed below. Each of the two DNA amplification reactions were carried out using a Perkin Elmer/Cetus DNA Thermal Cycler. The 40µl reaction mix consisted of 1X PCR Buffer (Cetus), 0.2mM each of the 4 dXTPs (Cetus), 50 pmoles of each primer oligonucleotide, 2ng of G-CSF template DNA (on a plasmid vector), and 1 unit of Taq polymerase (Cetus). The amplification process was carried out for 30 cycles. Each cycle consisted of 1 minute at 94°C, 2 minutes at 50°C, and 3 minutes at 72°C.

DNA amplification "A" used the oligonucleotides:

5' CCACTGGCGGTGATACTGAGC 3' (Seq.ID. No.63) and

5' AGCAGAAAGCTTTCCGGCAGAGAAGAAGCAGGA 3' (Seq.ID. No. 64)

DNA amplification "B" used the oligonucleotides:
5' GCCGCAAAGCTTTCTGCTGAAATGTCTGGAAGAGGTTCGTAAAATCCAGGGTGA 3'
(Seq.ID. No.65) and
5' CTGGAATGCAGAAGCAAATGCCGGCATAGCACCTTCAGTCGGTTGCAGAGCTGGTGCCA 3'
(Seq.ID. No.66)

From the 109 base pair double stranded DNA product obtained after DNA amplification "A", a 64 base pair XbaI to HindIII DNA fragment was cut and isolated that contained the DNA mutation Gln^{12}->Glu^{12}. From the 509 base pair double stranded DNA product obtained after DNA amplification "B", a 197 base pair HindIII to BsmI DNA fragment was cut and isolated that contained the DNA mutation Gln^{21}->Glu^{21}.

The "A" and "B" fragments were ligated together with a 4.8 kilo-base pair XbaI to BsmI DNA plasmid vector fragment. The ligation mix consisted of equal molar DNA restriction fragments, ligation buffer (25mM Tris-HCl pH 7.8, 10mM $MgCl_2$, 2mM DTT, 0.5mM rATP, and 100µg/ml BSA) and T4 DNA ligase and was incubated overnight at 14°C. The ligated DNA was then transformed into E. coli FM5 cells by electroporation using a Bio Rad Gene Pulsar apparatus (BioRad, Richmond, CA). A clone was isolated and the plasmid construct verified to contain the two mutations by DNA sequencing. This "intermediate" vector also contained a deletion of a 193 base pair BsmI to BsmI DNA fragment. The final plasmid vector was constructed by ligation and transformation (as described above) of DNA fragments obtained by cutting and isolating a 2 kilo-base pair SstI to BamHI DNA fragment from the intermediate vector, a 2.8 kbp SstI to EcoRI DNA fragment from the plasmid vector, and a 360 bp BamHI to EcoRI DNA fragment from the plasmid vector. The final

construct was verified by DNA sequencing the G-CSF gene. Cultures were grown, and the cells were harvested, and the G-CSF analogs were purified as set forth below.

As indicated above, any combination of mutagenesis techniques may be used to generate a G-CSF analog nucleic acid (and expression product) having one or more than one alteration. The two examples above, using M13-based mutagenesis and gene amplification-based mutagenesis, are illustrative.

E. Expression of G-CSF Analog DNA

The G-CSF analog DNAs were then placed into a plasmid vector and used to transform E. coli strain FM5 (ATCC#53911). The present G-CSF analog DNAs contained on plasmids and in bacterial host cells are available from the American Type Culture Collection, Rockville, MD, and the accession designations are indicated below.

One liter cultures were grown in broth containing 10g tryptone, 5g yeast extract and 5g NaCl) at 30°C until reaching a density at A^{600} of 0.5, at which point they were rapidly heated to 42°C. The flasks were allowed to continue shaking at for three hours.

Other prokaryotic or eukaryotic host cells may also be used, such as other bacterial cells, strains or species, mammalian cells in culture (COS, CHO or other types) insect cells or multicellular organs or organisms, or plant cells or multicellular organs or organisms, and a skilled practitioner will recognize the appropriate host. The present G-CSF analogs and related compositions may also be prepared synthetically, as, for example, by solid phase peptide

synthesis methods, or other chemical manufacturing techniques. Other cloning and expression systems will be apparent to those skilled in the art.

F. Purification of G-CSF Analog Protein

Cells were harvested by centrifugation (10,000 x G, 20 minutes, 4°C). The pellet (usually 5 grams) was resuspended in 30ml of 1mM DTT and passed three times through a French press cell at 10,000 psi. The broken cell suspension was centrifuged at 10,000g for 30 minutes, the supernatant removed, and the pellet resuspended in 30-40ml water. This was recentrifuged at 10,000 x G for 30 minutes, and this pellet was dissolved in 25ml of 2% Sarkosyl and 50mM Tris at pH 8. Copper sulfate was added to a concentration of 40μM, and the mixture was allowed to stir for at least 15 hours at 15-25°C. The mixture was then centrifuged at 20,000 x G for 30 minutes. The resultant solubilized protein mixture was diluted four-fold with 13.3 mM Tris, pH 7.7, after which was added approximately 20g Dowex™ (BioRad, Richmond, CA) equilibrated in 20mM Tris, pH 7.7. The mixture was stirred 90 minutes at room temperature and then the Dowex™ was filtered out. The supernatant was then applied to a DEAE-cellulose (Whatman DE-52) column equilibrated in 20mM Tris, pH 7.7. After loading and washing the column with the same buffer, the analogs were eluted with 20mM Tris/NaCl (between 35mM to 100mM depending on the analog, as indicated below), pH 7.7. For most of the analogs, the eluent from the DEAE column was adjusted to a pH of 5.4, with 50% acetic acid and diluted as necessary (to obtain the proper conductivity) with 5mM

sodium acetate pH 5.4. The solution was then loaded onto a CM-sepharose column equilibrated in 20mM sodium acetate, pH 5.4. The column was then washed with 20mM NaAc, pH 5.4 until the absorbance at 280nm was approximately zero. The G-CSF analog was then eluted with sodium acetate/NaCl in concentrations as described below in Table 4. The DEAE column eluents for those analogs not applied to the CM-sepharose column were dialyzed directly into 10mM NaAc, ph 4.0 buffer. The purified G-CSF analogs were then suitably isolated for in vitro analysis. The salt concentrations used for eluting the analogs varied, as noted above. Below, the salt concentrations for the DEAE cellulose column and for the CM-sepharose column are listed:

TABLE 4

Salt Concentrations

Analog	DEAE Cellulose	CM-Sepharose	SEQ.ID. NOS.
Lys^{17}->Arg^{17}	35mM	37.5mM	Seq.ID.No.67
Lys^{24}->Arg^{24}	35mM	37.5mM	Seq.ID.No.68
Lys^{35}->Arg^{35}	35mM	37.5mM	Seq.ID.No.69
Lys^{41}->Arg^{41}	35mM	37.5mM	Seq.ID.No.70
$Lys^{17,24,35}$->$Arg^{17,24,35}$	35mM	37.5mM	Seq.ID.No.71
$Lys^{17,35,41}$->$Arg^{17,35,41}$	35mM	37.5mM	Seq.ID.No.72
$Lys^{24,35,41}$->$Arg^{24,35,41}$	35mM	37.5mM	Seq.ID.No.73
$Lys^{17,24,35,41}$ ->$Arg^{17,24,35,41}$	35mM	37.5mM	Seq.ID.No.74
$Lys^{17,24,41}$->$Arg^{17,24,41}$	35mM	37.5mM	Seq.ID.No.75
Gln^{68}->Glu^{68}	60mM	37.5mM	Seq.ID.No.76
$Cys^{37,43}$->$Ser^{37,43}$	40mM	37.5mM	Seq.ID.No.77
Gln^{26}->Ala^{26}	40mM	40mM	Seq.ID.No.78
Gln^{174}->Ala^{174}	40mM	40mM	Seq.ID.No.79
Arg^{170}->Ala^{170}	40mM	40mM	Seq.ID.No.80
Arg^{167}->Ala^{167}	40mM	40mM	Seq.ID.No.81
Deletion 167*	N/A	N/A	Seq.ID.No.82
Lys^{41}->Ala^{41}	160mM	40mM	Seq.ID.No.83
His^{44}->Lys^{44}	40mM	60mM	Seq.ID.No.84
Glu^{47}->Ala^{47}	40mM	40mM	Seq.ID.No.85
Arg^{23}->Ala^{23}	40mM	40mM	Seq.ID.No.86
Lys^{24}->Ala^{24}	120mM	40mM	Seq.ID.No.87
Glu^{20}->Ala^{20}	40mM	60mM	Seq.ID.No.88
Asp^{28}->Ala^{28}	40mM	80mM	Seq.ID.No.89
Met^{127}->Glu^{127}	80mM	40mM	Seq.ID.No.90
Met^{138}->Glu^{138}	80mM	40mM	Seq.ID.No.91
Met^{127}->Leu^{127}	40mM	40mM	Seq.ID.No.92

Met^{138}->Leu^{138}	40mM	40mM	Seq.ID.No.93
Cys^{18}->Ala^{18}	40mM	37.5mM	Seq.ID.No.94
$Gln^{12,21}$->$Glu^{12,21}$	60mM	37.5mM	Seq.ID.No.95
$Gln^{12,21,68}$->$Glu^{12,21,68}$	60mM	37.5mM	Seq.ID.No.96
Glu^{20}->Ala^{20}; Ser^{13}->Gly^{13}	40mM	80mM	Seq.ID.No.97
$Met^{127,138}$->$Leu^{127,138}$	40mM	40mM	Seq.ID.No.98
Ser^{13}->Ala^{13}	40mM	40mM	Seq.ID.No.99
Lys^{17}->Ala^{17}	80mM	40mM	Seq.ID.No.100
Gln^{121}->Ala^{121}	40mM	60mM	Seq.ID.No.101
Gln^{21}->Ala^{21}	50mM	Gradient 0-150mM	Seq.ID.No.102
His^{44}->Ala^{44}**	40mM	N/A	Seq.ID.No.103
His^{53}->Ala^{53}**	50mM	N/A	Seq.ID.No.104
Asp^{110}->Ala^{110}**	40mM	N/A	Seq.ID.No.105
Asp^{113}->Ala^{113}**	40mM	N/A	Seq.ID.No.106
Thr^{117}->Ala^{117}**	50mM	N/A	Seq.ID.No.107
Asp^{28}->Ala^{28}; Asp^{110}-> Ala^{110}**	50mM	N/A	Seq.ID.No.108
Glu^{124}->Ala^{124}**	40mM	40mM	Seq.ID.No.109

* For Deletion 167, the data are unavailable.
** For these analogs, the DEAE cellulose column alone was use for purification.

The above purification methods are illustrative, and a skilled practitioner will recognize that other means are available for obtaining the present G-CSF analogs.

G. Biological Assays

Regardless of which methods were used to create the present G-CSF analogs, the analogs were subject to assays for biological activity. Tritiated thymidine assays were conducted to ascertain the degree of cell division. Other biological assays, however, may be used to ascertain the desired activity. Biological assays such as assaying for the ability to induce terminal differentiation in mouse WEHI-3B (D+) leukemic cell line, also provides indication of G-CSF activity. See Nicola, et al. *Blood* 54: 614-27 (1979). Other in vitro assays may be used to ascertain biological activity. See Nicola, *Ann.Rev. Biochem.* 58: 45-77 (1989). In general, the test for biological activity should provide analysis for the desired result, such as increase or decrease in biological

activity (as compared to non-altered G-CSF), different biological activity (as compared to non-altered G-CSF), receptor affinity analysis, or serum half-life analysis. The list is incomplete, and those skilled in the art will recognize other assays useful for testing for the desired end result.

The ^{3}H-thymidine assay was performed using standard methods. Bone marrow was obtained from sacrificed female Balb C mice. Bone marrow cells were briefly suspended, centrifuged, and resuspended in a growth medium. A 160µl aliquot containing approximately 10,000 cells was placed into each well of a 96 well micro-titer plate. Samples of the purified G-CSF analog(as prepared above) were added to each well, and incubated for 68 hours. Tritiated thymidine was added to the wells and allowed to incubate for five additional hours. After the five hour incubation time, the cells were harvested, filtered, and thoroughly rinsed. The filters were added to a vial containing scintillation fluid. The beta emissions were counted (LKB Betaplate scintillation counter). Standards and analogs were analyzed in triplicate, and samples which fell substantially above or below the standard curve were re-assayed with the proper dilution. The results reported here are the average of the triplicate analog data relative to the unaltered recombinant human G-CSF standard results.

H. HPLC Analysis

High pressure liquid chromatography was performed on purified samples of analog. Although peak position on a reverse phase HPLC column is not a definitive indication of structural similarity between

two proteins, analogs which have similar retention times may have the same type of hydrophobic interactions with the HPLC column as the non-altered molecule. This is one indication of an overall similar structure.

Samples of the analog and the non-altered recombinant human G-CSF were analyzed on a reverse phase (0.46 x 25cm) Vydac 214TP54 column (Separations Group, Inc. Hesperia, CA). The purified analog G-CSF samples were prepared in 20mM acetate and 40mM NaCl solution buffered at pH 5.2 to a final concentration of 0.1mg/ml to 5mg/ml, depending on how the analog performed in the column. Varying amounts (depending on the concentration) were loaded onto the HPLC column, which had been equilibrated with an aqueous solution containing 1% isopropanol, 52.8% acetonitrile, and 38% trifluoro acetate (TFA). The samples were subjected to a gradient of 0.86%/minute acetonitrile, and .002% TFA.

I. Results

Presented below are the results of the above biological assays and HPLC analysis. Biological activity is the average of triplicate data and reported as a percentage of the control standard (non-altered G-CSF). Relative HPLC peak position is the position of the analog G-CSF relative to the control standard (non-altered G-CSF) peak. The "+" or "-" symbols indicate whether the analog HPLC peak was in advance of or followed the control standard peak (in minutes). Not all of the variants had been analyzed for relative HPLC peak, and only those so analyzed are included below. Also presented are the American Type Culture Collection designations for E. coli host cells containing the

nucleic acids coding for the present analogs, as prepared above.

Table 5

Variant	Analog	Relative HPLC Peak	ATCC No.	% Normal G-CSF Activity	SEQ. ID. NOS.
1	Lys^{17}->Arg^{17}	N/A	69184	N/A	67
2	Lys^{24}->Arg^{24}	N/A	69185	N/A	68
3	Lys^{35}->Arg^{35}	N/A	69186	N/A	69
4	Lys^{41}->Arg^{41}	N/A	69187	N/A	70
5	$Lys^{17,24,35}$->$Arg^{17,24,35}$	N/A	69189	N/A	71
6	$Lys^{17,35,41}$->$Arg^{17,35,41}$	N/A	69192	N/A	72
7	$Lys^{24,35,41}$->$Arg^{24,35,41}$	N/A	69191	N/A	73
8	$Lys^{17,24,35,41}$->$Arg^{17,24,35,41}$	N/A	69193	N/A	74
9	$Lys^{17,24,41}$->$Arg^{17,24,41}$	N/A	69190	N/A	75
10	Gln^{68}->Glu^{68}	N/A	69196	N/A	76
11	$Cys^{37,43}$->$Ser^{37,43}$	N/A	69197	N/A	77
12	Gln^{26}->Ala^{26}	+.96	69201	51%	78
13	Gln^{174}->Ala^{174}	+.14	69202	100%	79
14	Arg^{170}->Ala^{170}	+.78	69203	100%	80
15	Arg^{167}->Ala^{167}	+.54	69204	110%	81
16	Deletion ^{167}	-.99	69207	N/A	82
17	Lys^{41}->Ala^{41}	+.25	69208	81%	83
18	His^{44}->Lys^{44}	-1.53	69212	70%	84
19	Glu^{47}->Ala^{47}	+.14	69205	0%	85
20	Arg^{23}->Ala^{23}	-.03	69206	31%	86
21	Lys^{24}->Ala^{24}	+1.95	69213	0%	87
22	Glu^{20}->Ala^{20}	-0.07	69211	0%	88
23	Asp^{28}->Ala^{28}	-.30	69210	147%	89
24	Met^{127}->Glu^{127}	N/A	69223	N/A	90
25	Met^{138}->Glu^{138}	N/A	69222	N/A	91
26	Met^{127}->Leu^{127}	N/A	69198	N/A	92
27	Met^{138}->Leu^{138}	N/A	69199	N/A	93
28	Cys^{18}->Ala^{18}	N/A	69188	N/A	94
29	$Gln^{12,21}$->$Glu^{12,21}$	N/A	69194	N/A	95
30	$Gln^{12,21,68}$->$Glu^{12,21,68}$	N/A	69195	N/A	96
31	Glu^{20}->Ala^{20} ; Ser^{13}->Gly^{13}	+1.74	69209	0%	97
32	$Met^{127,138}$->$Leu^{127,138}$	+1.43	69200	98%	98
33	Ser^{13}->Ala^{13}	0	69221	110%	99
34	Lys^{17}->Ala^{17}	+.50	69226	70%	100
35	Gln^{121}->Ala^{121}	+2.7	69225	100%	101
36	Gln^{21}->Ala^{21}	+0.63	69217	9.6%	102
37	His^{44}->Ala^{44}	+1.52	69215	10.8%	103
38	His^{53}->Ala^{53}	+0.99	69219	8.3%	104

39	Asp^{110}->Ala^{110}	+1.97	69216	29%	105
40	Asp^{113}->Ala^{113}	-0.34	69218	0%	106
41	Thr^{117}->Ala^{117}	+0.4	69214	9.7%	107
42	Asp^{28}->Ala^{28}; Asp^{110} Ala^{110}	+3.2	69220	20.6%	108
43	Glu^{124}->Ala^{124}	+0.16	69224	75%	109
44	Phe^{114}->Val^{114}, Thr^{117}-> Ala^{117}**	+0.53		0%	110

**This analog was apparently a result of an inadvertent error in the oligo which was used to prepare number 41, above (Thr^{117}-> Ala^{117}), and thus was prepared identically to the process used for that analog.
"N/A" indicates data which are not available.

1. Identification of Structure-Function Relationships

The first step used to design the present analogs was to determine what moieties are necessary for structural integrity of the G-CSF molecule. This was done at the amino acid residue level, although the atomic level is also available for analysis. Modification of the residues necessary for structural integrity results in change in the overall structure of the G-CSF molecule. This may or may not be desirable, depending on the analog one wishes to produce. The working examples here were designed to maintain the overall structural integrity of the G-CSF molecule, for the purpose of maintain G-CSF receptor binding of the analog to the G-CSF receptor (as used in this section below, the "G-CSF receptor" refers to the natural G-CSF receptor, found on hematopoietic cells). It was assumed, and confirmed by the studies presented here, that G-CSF receptor binding is a necessary step for at least one biological activity, as determined by the above biological assays.

As can be seen from the figures, G-CSF (here, recombinant human met-G-CSF) is an antiparallel 4-alpha

helical bundle with a left-handed twist, and with overall dimensions of 45Å x 30Å x 24Å . The four helices within the bundle are referred to as helices A, B, C and D, and their connecting loops are known as the AB, BC and CD loops. The helix crossing angles range from -167.5° to -159.4°. Helices A, B, and C are straight, whereas helix D contains two kinds of structural characteristics, at Gly^{150} and Ser^{160} (of the recombinant human met-G-CSF). Overall, the G-CSF molecule is a bundle of four helices, connected in series by external loops. This structural information was then correlated with known functional information. It was known that residues (including methionine at position 1) 47, 23, 24, 20, 21, 44, 53, 113, 110, 28 and 114 may be modified, and the effect on biological activity would be substantial.

The majority of single mutations which lowered biological activity were centered around two regions of G-CSF that are separated by 30Å, and are located on different faces of the four helix bundle. One region involves interactions between the A helix and the D helix. This is further confirmed by the presence of salt bridges in the non-altered molecule as follows:

Atom	Helix	Atom	Helix	Distance
Arg^{170} N1	D	Tyr^{166} OH	A	3.3
Tyr^{166} OH	D	Arg^{23} N2	A	3.3
Glu^{163} OE1	D	Arg^{23} N1	A	2.8
Arg^{23} N1	A	Gln^{26} OE1	A	3.1
Gln^{159} NE2	D	Gln^{26} O	A	3.3

Distances reported here were for molecule A, as indicated in FIGURE 5 (wherein three G-CSF molecules

crystallized together and were designated as A, B, and C). As can be seen, there is a web of salt bridges between helix A and helix D, which act to stabilize the helix A structure, and therefore affect the overall structure of the G-CSF molecule.

The area centering around residues Glu^{20}, Arg^{23} and Lys^{24} are found on the hydrophilic face of the A helix (residues 20-37). Substitution of the residues with the non-charged alanine residue at positions 20 and 23 resulted in similar HPLC retention times, indicating similarity in structure. Alteration of these sites altered the biological activity (as indicated by the present assays). Substitution at Lys^{24} altered biological activity, but did not result in a similar HPLC retention time as the other two alterations.

The second site at which alteration lowered biological activity involves the AB helix. Changing glutamine at position 47 to alanine (analog no. 19, above) reduced biological activity (in the thymidine uptake assay) to zero. The AB helix is predominantly hydrophobic, except at the amino and carboxy termini; it contains one turn of a 3^{10} helix. There are two histadines at each termini (His^{44} and His^{56}) and an additional glutamate at residue 46 which has the potential to form a salt bridge to His^{44}. The fourier transformed infra red spectrographic analysis (FTIR) of the analog suggests this analog is structurally similar to the non-altered recombinant G-CSF molecule. Further testing showed that this analog would not crystallize under the same conditions as the non-altered recombinant molecule.

Alterations at the carboxy terminus (Gln^{174}, Arg^{167} and Arg^{170}) had little effect on biological activity. In contrast, deletion of the last eight residues (167-175) lowered biological activity. These results may indicate that the deletion destabilizes the overall structure which prevents the mutant from proper binding to the G-CSF receptor (and thus initiating signal transduction).

Generally, for the G-CSF internal core -- the internal four helix bundle lacking the external loops -- the hydrophobic internal residues are essential for structural integrity. For example, in helix A, the internal hydrophobic residues are (with methionine being position 1) Phe^{14}, Cys^{18}, Val^{22}, Ile^{25}, Ile^{32} and Leu^{36}. Generally, for the G-CSF internal core -- the internal four helix bundle lacking the external loops - the hydrophobic internal residues are essential for structural integrity. For example, in helix A, the internal hydrophobic residues are (with methionine being position 1 as in FIGURE 1, Seq.ID. No.2) Phe^{14}, Cys^{18}, Val^{22}, Ile^{25}, Ile^{32} and Leu^{36}. The other hydrophobic residues (again with the met at position 1) are: helix B , Ala^{72}, Leu^{76}, Leu^{79}, Leu^{83}, Tyr^{86}, Leu^{90} Leu^{93}; helix C, Leu^{104}, Leu^{107}, Val^{111}, Ala^{114}, Ile^{118}, Met^{122}; and helix D, Val^{154}, Val^{158}, Phe^{161}, Val^{164}, Val^{168}, Leu^{172}.

The above biological activity data, from the presently prepared G-CSF analogs, demonstrate that modification of the external loops interfere least with G-CSF overall structure. Preferred loops for analog preparation are the AB loop and the CD loop. The loops are relatively flexible structures as compared to the helices. The loops may contribute to the proteolysis

of the molecule. G-CSF is relatively fast acting in vivo as the purpose the molecule serves is to generate a response to a biological challenge, i.e., selectively stimulate neutrophils. The G-CSF turnover rate is also relatively fast. The flexibility of the loops may provide a "handle" for proteases to attach to the molecule to inactivate the molecule. Modification of the loops to prevent protease degradation, yet have (via retention of the overall structure of non-modified G-CSF) no loss in biological activity may be accomplished.

This phenomenon is probably not limited to the G-CSF molecule but may also be common to the other molecules with known similar overall structures, as presented in FIGURE 2. Alteration of the external loop of, for example hGH, Interferon β, IL-2, GM-CSF and IL-4 may provide the least change to the overall structure. The external loops on the GM-CSF molecule are not as flexible as those found on the G-CSF molecule, and this may indicate a longer serum life, consistent with the broader biological activity of GM-CSF. Thus, the external loops of GM-CSF may be modified by releasing the external loops from the beta-sheet structure, which may make the loops more flexible (similar to those G-CSF) and therefore make the molecule more susceptible to protease degradation (and thus increase the turnover rate).

Alteration of these external loops may be effected by stabilizing the loops by connection to one or more of the internal helices. Connecting means are known to those in the art, such as the formation of a beta sheet, salt bridge, disulfide bonding or hydrophobic interactions, and other means are

available. Also, deletion of one or more moieties, such as one or more amino acid residues or portions thereof, to prepare an abbreviated molecule and thus eliminate certain portions of the external loops may be effected.

Thus, by alteration of the external loops, preferably the AB loop (amino acids 58-72 of r-hu-met G-CSF) or the CD loop (amino acids 119 to 145 of r-hu-met-G-CSF), and less preferably the amino terminus (amino acids 1-10), one may therefore modify the biological function without elimination of G-CSF G-CSF receptor binding. For example, one may: (1) increase half-life (or prepare an oral dosage form, for example) of the G-CSF molecule by, for example, decreasing the ability of proteases to act on the G-CSF molecule or adding chemical modifications to the G-CSF molecule, such as one or more polyethylene glycol molecules or enteric coatings for oral formulation which would act to change some characteristic of the G-CSF molecule as described above, such as increasing serum or other half-life or decreasing antigenicity; (2) prepare a hybrid molecule, such as combining G-CSF with part or all of another protein such as another cytokine or another protein which effects signal transduction via entry through the cell through a G-CSF G-CSF receptor transport mechanism; or (3) increase the biological activity as in, for example, the ability to selectively stimulate neutrophils (as compared to a non-modified G-CSF molecule). This list is not limited to the above exemplars.

Another aspect observed from the above data is that stabilizing surface interactions may affect biological activity. This is apparent from comparing

analogs 23 and 40. Analog 23 contains a substitution of the charged asparagine residue at position 28 for the neutrally-charged alanine residue in that position, and such substitution resulted in a 50% increase in the biological activity (as measured by the disclosed thymidine uptake assays). The asparagine residue at position 28 has a surface interaction with the asparagine residue at position 113; both residues being negatively charged, there is a certain amount of instability (due to the repelling of like charged moieties). When, however the asparagine at position 113 is replaced with the neutrally-charged alanine, the biological activity drops to zero (in the present assay system). This indicates that the asparagine at position 113 is critical to biological activity, and elimination of the asparagine at position 28 serves to increase the effect that asparagine at position 113 possesses.

The domains required for G-CSF receptor binding were also determined based on the above analogs prepared and the G-CSF structure. The G-CSF receptor binding domain is located at residues (with methionine being position 1) 11-57 (between the A and AB helix) and 100-118 (between the B and C helices). One may also prepare abbreviated molecules capable of binding to a G-CSF receptor and initiate signal transduction for selectively stimulating neutrophils by changing the external loop structure and having the receptor binding domains remain intact.

Residues essential for biological activity and presumably G-CSF receptor binding or signal transduction have been identified. Two distinct sites are located on two different regions of the secondary

structure. What is here called "Site A" is located on a helix which is constrained by salt bridge contacts between two other members of the helical bundle. The second site, "Site B" is located on a relatively more flexible helix, AB. The AB helix is potentially more sensitive to local pH changes because of the type and position of the residues at the carboxy and amino termini. The functional importance of this flexible helix may be important in a conformationally induced fit when binding to the G-CSF receptor. Additionally, the extended portion of the D helix is also indicated to be a G-CSF receptor binding domain, as ascertained by direct mutational and indirect comparative protein structure analysis. Deletion of the carboxy terminal end of r-hu-met-G-CSF reduces activity as it does for hGH, see, Cunningham et al. *Science* 244: 1081-1084 (1989). Cytokines which have similar structures, such as IL-6 and GM-CSF with predicted similar topology also center their biological activity along the carboxy end of the D helix, see Bazan *Immunology Today* 11: 350-354 (1990).

A comparison of the structures and the positions of G-CSF receptor binding determinants between G-CSF and hGH suggests both molecules have similar means of signal transduction. Two separate G-CSF receptor binding sites have been identified for hGH De Vos et al. *Science* 255: 306-32 (1991). One of these binding sites (called "Site I") is formed by residues on the exposed faces of hGH's helix 1, the connection region between helix 1 and 2, and helix 4. The second binding site (called "Site II") is formed by surface residues of helix 1 and helix 3.

The G-CSF receptor binding determinates identified for G-CSF are located in the same relative positions as those identified for hGH. The G-CSF receptor binding site located in the connecting region between helix A and B on the AB helix (Site A) is similar in position to that reported for a small piece of helix (residues 38-47) of hGH. A single point mutation in the AB helix of G-CSF significantly reduces biological activity (as ascertained in the present assays), indicating the role in a G-CSF receptor-ligand interface. Binding of the G-CSF receptor may destabilize the 3^{10} helical nature of this region and induce a conformation change improving the binding energy of the ligand/G-CSF receptor complex.

In the hGH receptor complex, the first helix of the bundle donates residues to both of the binding sites required to dimerize the hGH receptor. Mutational analysis of the corresponding helix of G-CSF (helix A) has identified three residues which are required for biological activity. Of these three residues, Glu^{20} and Arg^{24} lie on one face of the helical bundle towards helix C, whereas the side chain of Arg^{23} (in two of the three molecules in the asymmetric unit) points to the face of the bundle towards helix D. The position of side chains of these biologically important residues indicates that similar to hGH, G-CSF may have a second G-CSF receptor binding site along the interface between helix A and helix C. In contrast with the hGH molecule, the amino terminus of G-CSF has a limited biological role as deletion of the first 11 residues has little effect on the biological activity.

As indicated above (see FIGURE 2, for example), G-CSF has a topological similarity with other

cytokines. A correlation of the structure with previous biochemical studies, mutational analysis and direct comparison of specific residues of the hGH receptor complex indicates that G-CSF has two receptor binding sites. Site A lies along the interface of the A and D helices and includes residues in the small AB helix. Site B also includes residues in the A helix but lies along the interface between helices A and C. The conservation of structure and relative positions of biologically important residues between G-CSF and hGH is one indication of a common method of signal transduction in that the receptor is bound in two places. It is therefore found that G-CSF analogs possessing altered G-CSF receptor binding domains may be prepared by alteration at either of the G-CSF receptor binding sites (residues 20-57 and 145-175).

Knowledge of the three dimensional structure and correlation of the composition of G-CSF protein makes possible a systematic, rational method for preparing G-CSF analogs. The above working examples have demonstrated that the limitations of the size and polarity of the side chains within the core of the structure dictate how much change the molecule can tolerate before the overall structure is changed.